

Vocabulary Learning and Instruction

Volume 7, Number 1,

December 2018

doi: <http://dx.doi.org/10.7820/vli.v07.1.2187-2759>

VLI Editorial Team

Editorial Board: Jeffrey Stewart, Raymond Stubbe, Aaron Gibson

Reviewers: Paul Meara, Norbert Schmitt, John A. Read, Stuart Webb, John P. Racine, Tomoko Ishii, Tim Stoeckel, Phil Bennett, Dale Brown, Jon Clenton, Stuart McLean, Peter Thwaites, Tatsuya Nakata, Kiwamu Kasahara, Masumi Kojima, James Rogers, Yuko Hoshino, Vivienne Rogers, Yu Kanazawa, Stephen Cutler, Ian Munby, Mark Maby, Brandon Kramer

Copy Editors: Alex Cameron, Andrew Gallacher, Peter Harold, Mark Howarth, Linda Joyce, Tim Pritchard, Andrew Thompson, Alonzo Williams, Ian Mumby, Pete Westbrook, Melanie González, Stephen Cutler, Michael McGuire, Shusaku Kida, Jenifer Larson-Hall, Brent Culligan, Tim Stoeckel, Shin Ishikawa, Magda Kitano.

The Editorial Team expresses a sincere thank you to Mana Ikawa, who designed the cover for the print version of *VLI*.

Copyright © 2018 *Vocabulary Learning and Instruction*, ISSN: Online 2187-2759; Print 2187-2767. All articles are copyrighted by their respective authors.

Vocabulary Learning and Instruction

Volume 7, Number 1, December 2018

doi: <http://dx.doi.org/10.7820/vli.v07.1.2187-2759>

Table of Contents

Articles	Page
Letter from the Editor <i>Raymond Stubbe</i>	iv
Regular Articles	
The Indirect Effects of Testing: Can Poor Performance in a Vocabulary Quiz Lead to Long-Term L2 Vocabulary Retention? <i>Kohei Kanayama and Kiwamu Kasahara</i>	1
Vocabulary Learning Strategies and Vocabulary Size: Insights from Educational Level and Learner Styles <i>Alaa Alahmadi, Christopher Shank, and Anouschka Foltz</i>	14
An Experimental Investigation of HAM, a Novel Mnemonic Technique for Learning L2 Homonyms and Homophones <i>Charles M. Mueller</i>	35
Examining the Word Family through Word Lists <i>Dale Brown</i>	51

Letter from the Editor

Dear Readers,

It is my great pleasure to offer you the November 2018 issue of *Vocabulary Learning and Instruction (VLI)*. Our first paper will “examine whether restudying immediately after the failure in the test is useful for long-term retention” (p. 1). The next article “investigates the effect of different vocabulary learning strategies (VLS) as well as different learner styles on vocabulary size in Saudi Arabic-speaking students in higher education” (p. 14). The third “presents a novel approach called the Homonym/Homophone Association Method (HAM) ... {which} overcomes some of the drawbacks of the keyword method by associating meanings of L2 homonyms or homophones, one known by the learner and one unknown” (p. 35). The final article’s “aim is to conduct a thorough examination of a set of word-family-based word lists in order to provide a detailed characterization of such lists” (p. 54). This analysis covered the first five bands of *Nation’s (2006a) British National Corpus-based word lists*, and a link to the supporting spreadsheet *Occurrences of affixes in the 1K-5K word families* is available. It is anticipated that this link will be well used by our readers.

As a reminder, VLI is an open-access international journal that provides a peer reviewed forum for original research related to vocabulary acquisition, instruction, and assessment. Submissions are encouraged from researchers and practitioners in both EFL and ESL contexts.

Please enjoy this issue,
Raymond Stubbe,
Editor, VLI

The Indirect Effects of Testing: Can Poor Performance in a Vocabulary Quiz Lead to Long-Term L2 Vocabulary Retention?

Kohei Kanayama^a and Kiwamu Kasahara^b

^a*Sapporo Otani High School*; ^b*Hokkaido University of Education*

doi: <https://doi.org/10.7820/vli.v07.1.kanayama.kasahara>

Abstract

Taking a test on learned items enhances long-term retention of these items. However, it is believed that good performance in a test contributes to subsequent high retention of the tested items while poor performance does not. Recent studies have sought to find the optimal way to make up for this poor performance, and have indicated that giving the subsequent learning session soon after the test is one such way. This study is different from previous studies in that we used L1–L2 word pairs to examine whether restudying immediately after the failure in the test is useful for long-term retention. First, in the initial study session, all the participants ($n = 52$) were shown and asked to remember 20 English and Japanese word pairs (e.g., deceit:詐欺). A week later, Group A took the first test session (Initial Test) before the restudy session. On the contrary, Group B took the restudy session before the Initial Test. An hour after this session, both groups took Posttest 1. Then, Posttest 2 was conducted a week after Posttest 1. The results showed that Group A had significantly lower scores than Group B in the Initial Test (2% vs. 55%). However, the results were reversed in Posttest 1 (84.2% vs. 53.2%) and Posttest 2 (55% vs. 43.5%). This study found that a restudy session soon after poor performance in the Initial Test enhanced long-term L2 vocabulary retention because learners benefited from the indirect effects of testing. Thus, English teachers should take such effects into consideration when organizing vocabulary quizzes and restudy sessions.

1 Introduction

There is a great deal of evidence that taking a test on learned items can strengthen long-term retention of these items. This phenomenon has been widely demonstrated as the *testing effect* (Barcroft, 2007; Carpenter, Pashler, & Vul, 2006; Halamish & Bjork, 2011; Karpicke & Roediger, 2007; Meyer & Logan, 2013). Previous studies have demonstrated that tests have positive effects on any educationally relevant materials such as fact learning (Agarwal, Finley, Rose, & Roediger, 2017; Butler & Roediger, 2007; Carpenter, Pashler, Wixted, & Vul, 2008; Roediger, Agarwal, McDaniel, & McDermott, 2011a), learning from reading passages (McDaniel, Anderson, Derbish, & Morrisette, 2007a; Roediger & Karpicke, 2006b), picture-L2 word pairs (Barcroft, 2007), cue–target word pair learning (Carpenter et al., 2006; Cull, 2000; Halamish & Bjork, 2011; Karpicke & Roediger,

2007; Roediger & Marsh, 2005), and L1–L2 pair learning (Kanayama & Kasahara, 2015). Thus, a test itself can be a powerful tool to improve students' performance in a later test. However, the testing effect has been ignored in actual school settings (Carpenter et al., 2008; Karpicke & Grimaldi, 2012; McDaniel et al., 2007). In other words, teachers have generally viewed tests as just tools to evaluate students' current learning achievements (Butler & Roediger, 2007), to assign grades to the students (Roediger, Putnam, & Smith, 2011b; Vojdanoska, Cranney, & Newell, 2010), or just as means to assess what has been learned (Blunt & Karpicke, 2014; Karpicke & Roediger, 2007). Few teachers have seen the test as a tool to improve students' performance. However, some researchers emphasize on the importance of the test by stating that “tests should be used frequently in educational contexts not merely to assess learning, as is the standard practice, but to promote it” (Carpenter et al., 2008, p. 438), and “testing to enhance learning should be seriously considered in pedagogical theory and practice” (McDaniel et al., 2007, p. 510). Therefore, the importance of the tests as a tool to enhance students' long-term retention should be reviewed.

2 Literature Review

2.1 *The Direct Effects of Testing on Long-Term Retention*

Tests have a positive effect on learning. Roediger et al. (2011) distinguish two types of testing effect: direct and indirect effects of testing. The direct effect means that taking a test itself has a positive effect on learning (Carrier & Pashler, 1992; Roediger & Karpicke, 2006a). Learners can benefit greatly from just taking a test on learned items. One of the reasons is that tests always include *retrieval practice* (Nation, 2013). This is a process in which learners try to search for previously learned items in their memory. In case of vocabulary meaning retrieval, test-takers are often asked to retrieve L1 equivalents of target words by seeing the L2 forms of the targets. In the case of vocabulary form retrieval, they are often asked to retrieve L2 forms by seeing their L1 equivalents. The act of retrieving the items itself enhances the long-term retention of them (Karpicke, 2012) because it requires the learners to make a mental effort to take out something from their memory with the help of cues, which can strengthen connections between the cues and the targets. These reinforced connections, or deeper traces in memory, can contribute to long-term retention (Craik & Tulving, 1975). If the items are successfully retrieved from memory, they will be established more strongly in learners' minds. As a result, the learners can put the successfully retrieved items into their long-term memory (Tullis, Finley, & Benjamin, 2013).

Halamish and Bjork (2011) found a positive effect of retrieval practice on vocabulary learning. In their study, a retrieval group and a non-retrieval group were presented with L1 cue–target word pairs in random order for 3 seconds each (e.g., *RENT: HOUSE*). After the first presentation, the non-retrieval group restudied each word pair for 6 seconds in the other two cycles. On the contrary, the retrieval group looked at each cue word and a part of the target word for 6 seconds (e.g., *RENT: __SE*) in the remaining two cycles. An immediate test asking both groups to recall the targets for the cue words (e.g., *RENT: ____?*) found that the retrieval

group recalled more of the target words than the non-retrieval group (39% vs. 13%). Kanayama and Kasahara (2015) supported the findings of Halamish and Bjork, investigating the effects of word retrieval on L1–L2 pair learning (e.g., *ligament*: 靭帯). The retrieval group, who had a chance to retrieve the Japanese meanings of the target English words (e.g., *ligament*: _____?), outperformed the non-retrieval group in a test conducted a week later (31.2% vs. 22.1%).

The direct effects of testing can also be seen in passage recall. For instance, in a study by Roediger and Karpicke (2006b, Experiment 1), the researchers compared a group that took a test after the first study session with a group that had the restudy session after the first study session. The former and latter groups were called the ST and the SS groups, respectively (S stands for one study session and T stands for one test trial). The study session asked the participants to read the target English passages. In the test trial, participants were given a blank sheet of paper and asked to retrieve as much information from the passages as possible, but no correct answers were given after the test. The results showed that the ST group was superior in 1-week retention to the SS group (56% vs. 42%).

Roediger and Karpicke (2006b, Experiment 2) replicated their first experiment. The procedure was the same as for Experiment 1, except that they compared the SSSS, the SSST, and the STTT groups. The SSSS group was asked to take a study session four times in a row. Meanwhile, The SSST took a test session after three consecutive study sessions. The STTT group took a study session and three test sessions consecutively. As with Experiment 1, no feedback was given after the test. The delayed test conducted a week later found that the STTT group had a better score than the SSST group, and that the SSST group outperformed the SSSS group (61% vs. 56% vs. 40%). Roediger and Karpicke found that the more retrieval sessions the participants had during the learning phases, the better retention of the learned items they showed. These studies (Halamish & Bjork, 2011; Kanayama & Kasahara, 2015; Roediger & Karpicke, 2006b) demonstrated that learners can benefit greatly from just retrieving the learned information from their memory.

2.2 The Indirect Effects of Testing on Learning

The indirect effects of testing are another type of benefits which learners can gain through tests (Roediger & Karpicke, 2006a): test results can help test-takers to focus on items they have not mastered yet. Tests inform the students of what items they have or have not acquired (Brame & Biel, 2015; Carpenter et al., 2008; Roediger et al., 2011). Based on this experience, when they have a restudy session after the test, they are able to spend more time remembering the items which they have not mastered rather than those they have already acquired (Karpicke & Grimaldi, 2012; Soderstrom & Bjork, 2014; Son & Kornell, 2008).

Moreover, tests can prevent the learners from being overconfident (Roediger et al., 2011). Learners who have not taken any tests tend to become more confident about a later test (Brown, Roediger, & McDaniel, 2014). On the contrary, those who have taken test sessions are able to evaluate their actual learning achievement properly. The second experiment presented in Roediger and Karpicke (2006b) provides some evidence for this. After all the learning sessions, all the participants were asked to rate how well they could retrieve the learned information from the

passage they read, using a 7-point scale (1 = not very well; 7 = very well). The results revealed that the SSSS group gave higher scores ($M = 4.8$) than the SSST group ($M = 4.2$). The least confident group was the STTT group ($M = 4.0$). A week later, all the participants returned to the laboratory and took the delayed test. As mentioned above, the actual score in the delayed test showed that the STTT group was better than the SSST and the SSSS groups (61% vs. 56% vs. 40%). Figure 1 presents the participants' judgment of their learning and the actual scores of each group. The participants in the SSSS group felt that they had already achieved their learning goal. Hence, they may not have tried hard in the remaining study sessions. On the contrary, the tests prevented the participants in the STTT group from being overconfident because they understood that they had not achieved their learning goal yet. This is reliable evidence that the SSSS group, which had multiple study sessions, became overconfident about their future performance, and that the STTT group, which took the test trials, was able to avoid becoming overconfident.

In addition, tests include a beneficial by-product of indirect effects of testing. That is, if students take a test frequently, they can acquire a habit of preparing for tests. As a result, they tend to study more often and spontaneously (Putnam, Nestojko, & Roediger, 2016). Moreover, tests can help teachers allocate enough time to teach and review what most students do not understand (Roediger et al., 2011). In short, teachers as well as their students can understand actual learning achievements through tests.

A study by Karpicke and Roediger (2007) revealed that learners who gained benefits from the indirect effects of testing showed better recall performance on a delayed test than those who did not benefit from the indirect effects. In their Experiment 1, they had their participants learn cue–target word pairs in standard (STST), repeated study (SSST), or repeated test (STTT) groups. Again, no feedback (correct answers) were given after the test. The STTT group was expected to show superiority in long-term retention compared with other groups because the long-term benefits increased with the number of retrieval sessions given (Roediger & Karpicke, 2006b). Nevertheless, the test conducted at the end of all the four learning sessions found that the STST group showed superior to the STTT and the SSST (88% vs. 81% vs. 78%). Moreover, the delayed test carried out a week later showed that the STST group had a better score than the STTT and the SSST groups (68% vs. 64% vs. 57%). In addition, their Experiment 2 also compared

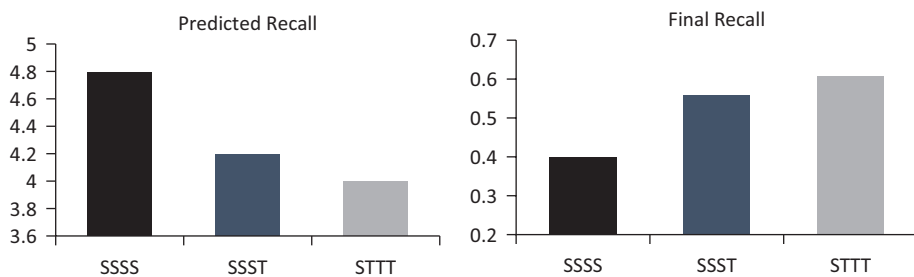


Figure 1. Predicted Scores and Final Performance on a 1-Week Delayed Test of the SSSS, SSST, and STTT Groups. Data Are Adapted From Roediger and Karpicke (2006b).

participants who learned in the STST with the SSTT groups. The results replicated their Experiment 1; the STST group showed higher retention rates than the SSTT group a week later (44% vs. 36%), indicating that the indirect effects of testing have long-term benefits.

A conceivable reason why the STST group showed better performance than the other groups (SSTT, SSST, and STTT) is that only the STST group gained more benefits from the indirect effects of testing because they had a chance to take a restudy session after a test. The first test helped the participants in the STST group notice what items they had not acquired, and helped them to evaluate themselves better. Based on this experience, they focused on the words which they had not mastered in the restudy session. Finally, they successfully recalled the items in the second test session. They benefited from both the direct and indirect effects of testing. On the contrary, the other three groups (SSTT, SSST, and STTT) benefited from the direct effects of testing (retrieval effect), but did not benefit from the indirect effects of testing because the three groups had no chance to take any restudy sessions after the tests.

Other past studies have also demonstrated that the indirect effects of testing exist on some educationally relevant materials such as learning from reading passages (Richland, Kornell, & Kao, 2009), cue–target word learning (Hays, Kornell, & Bjork, 2013; Kornell, Hays, & Bjork, 2009), and L1–L2 pair learning (Arnold & McDermott, 2013).

2.3 Previous Studies on Effects of Poor Performance in a Test on Learning

One of the reasons why indirect effects exist is that learners can narrow the items they should focus on in the subsequent study session after the test. However, it is still unclear whether the learners could gain much benefit from the indirect effects of testing even after they answered most items incorrectly. If most of the items were not recalled, such narrowing may be impossible. Therefore, it is worth investigating whether indirect effects exist even in the case where learners had a score very close to zero.

Some studies have tried to examine whether learners can benefit from the indirect effects of testing after poor performance in a test. However, these studies had some limitations. One of these studies was by Richland et al. (2009, Experiment 4). In their study, participants took the first test session in 2 minutes, where they were asked to answer the five questions relating to a short passage. Because they had not read the passage before answering the five questions, they answered nothing. After they experienced this failure, the participants took a study session in which they were asked to read a short essay (about cerebral achromatopsia) in 8 minutes. A week later, a posttest was conducted, where they were asked to answer 10 questions about the essay they read. Five out of the ten questions were in the first test session. The results showed that the five tested items were remembered better than the five untested items (55% vs. 42%). Richland et al. concluded that the learners were able to benefit from the indirect effects of testing after they showed poor performance in the previous test.

Kornell et al. (2009, Experiment 5) also had the same limitations as the study of Richland et al. In the study of Kornell et al., they had their participants remember cue–target word pairs (e.g., *train: caboose*). Participants in the read-only condition were given each pair for 13 seconds. On the contrary, those in the test condition were shown a cue word alone at first (e.g., *train: _____?*) for 8 seconds, followed by the cue and target for 5 seconds (e.g., *train: caboose*). Because the participants in the test condition had learned no word pairs, they could answer nothing in the test session. A day after the session, all the participants took the posttest, in which they were asked to recall the correct answer (e.g., *train: _____?*). It was found that the test group had a better score than the read-only condition (47% vs. 35%).

The limitation of the two studies is that there were some doubts that the participants truly showed poor performance in the test. Poor performance in a test refers to a situation in which learners failed to recall most of the learned items on tests although they had learned them before. In this respect, the participants in Richlands' and Kornells' study did not experience poor performance because they had learned no target items before the first study session. Rather, the participants in both studies benefited from the *pretesting effect*, the idea that testing before studying helps learners focus on what will be tested later. In summary, there have been some studies which demonstrated the indirect effects of testing on long-term retention, but no previous studies have investigated whether learners are able to benefit from the indirect effects of testing after their poor performance in a test.

2.4 Purposes of This Study

In order to deal with the issue mentioned above, this study compared the effects of the S-ST and S-TS conditions. (The hyphen means an interval of 1 week between the first learning and the first review session.) Learners who take the first test trial a week after the first study session are likely to forget almost all of the learned words. It is possible that this unsuccessful experience prevents them from gaining benefits from the indirect effects. Therefore, it is worth investigating whether the indirect effects of testing can function well after learners cannot recall most of the learned words. By examining the effect of the S-TS and the S-ST conditions, this study became original in that it was able to investigate whether the indirect effects of testing are useful in L1–L2 pair learning even after the learners have experienced poor performance in a test. The research question of this study is as follows:

RQ. Can the indirect effects of testing after poor performance in a test promote vocabulary learning?

3 Method

3.1 Participants

The participants in this study were 85 first-year Japanese university EFL students, and they were divided into Group A (S-TS) and Group B (S-ST).

The data from 33 participants who did not take part in one or more sessions were excluded from the analyses. Thus, the following analyses were based on 52 participants. Group A consisted of 23 students, whereas there were 29 students in Group B. All the participants had studied English in Japan for at least 6 years.

3.2 Materials

This study used 20 low-frequency English words and their Japanese equivalent pairs. These English words comprised 10 nouns and 10 verbs, which are not listed in *JACET 8000* (JACET, 2003). *JACET 8000* includes about 8000 basic English words in a frequency order which Japanese learners of English are recommended to remember. We assumed that none of the words would be known by all the participants. Each English target word and its Japanese translation are shown in Table 1.

3.3 Procedure

All the study sessions and the tests were conducted in English lessons regularly held at the university. The participants were assigned to one of the two learning conditions: the S-TS or the S-ST groups. First, both groups were asked to remember the 20 target word pairs, which were displayed on PowerPoint slides. Each word pair was presented on a screen in front of the participants. Each group had opportunities to encounter each word pair three times. In the first cycle, the participants looked at a target word and its L1 translation at the same time for 6 seconds (e.g., ligament:靭帯). In the remaining two cycles, each word pair was displayed in the same way as the first cycle for 4 seconds. The total learning time was 4 minutes and 40 seconds for both groups.

A week later, the S-TS group took the Initial Test. Immediately after the Initial Test, they took the restudying session in the same way as the first one was conducted. On the contrary, the S-ST group took the restudying session before the Initial Test. Both groups were given Posttest 1 an hour after their respective last sessions. The participants took Posttest 2 a week after Posttest 1. All the tests

Table 1. Target Nouns and Verbs

English	Japanese	English	Japanese
mutineer	反逆者	gnaw	～をかじる
lemur	キツネザル	smuggle	～を密輸する
ligament	靭帯	sterilize	～を消毒する
encroachment	侵略	sham	(病気などの)ふりをする
adhesive	接着剤	impute	人のせいにする
ointment	化粧用クリーム	belittle	～をけなす
deceit	詐欺	assail	激しく非難する
palliative	緩和剤	contort	～をねじる
janitor	用務員	foray	襲撃する
knack	こつ	immerse	(液体などに)ひたす

(Initial Test, Posttest 1, and Posttest 2) required both groups to recall all the 20 Japanese meanings for the English target words.

3.4 Scoring

Two points were awarded for each correct answer. When a student gave a correct answer but in a wrong part of speech, only one point was given. For example, in the case of the word “sterilize,” the correct answer was *sakkin-suru*, whereas one point was given for *sakkin*. The maximum score for each test was 40 points (2 points \times 20 words).

4 Results

Table 2 presents the mean scores and standard deviations of each test for Groups A and B. Figure 2 shows the mean scores in each test for Groups A and B. An analysis through a 2 (Group: S-TS and S-ST) \times 3 (Test: Initial Test, Posttest 1, and Posttest 2) mixed ANOVA revealed that there was a significant interaction between the two factors, Group and Test, $F(1, 50) = 263.8$, $p < 0.001$, $\eta^2 = 0.84$ (large effect size). In addition, a significant main effect of Test was also observed, $F(1, 50) = 223.4$, $p < 0.001$, $\eta^2 = 0.81$ (large effect size). However, there was no significant main effect of Group, $F(1, 50) = 0.53$, $p = 0.467$, $\eta^2 = 0.01$.

Table 2. Mean Scores and Standard Deviations of Each Test for Both Groups A and B (Full Mark = 40)

	Group A (S-TS)		Group B (S-ST)	
	Mean	SD	Mean	SD
Initial Test	0.82	1.85	22.0	8.78
Posttest 1	33.7	6.24	21.3	8.54
Posttest 2	22.0	8.24	17.4	7.89

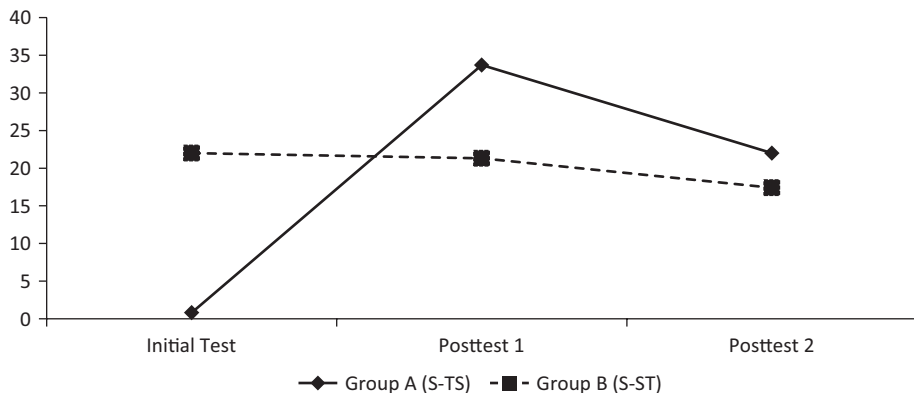


Figure 2. The Mean Scores of Each Test in Groups A and B.

Furthermore, there was a significant simple main effect between the two groups in the Initial Test, $F(1, 50) = 99.1, p < 0.001$; in Posttest 1, $F(1, 50) = 33.7, p < 0.001$; and in Posttest 2, $F(1, 50) = 4.6, p = 0.033$. This means that the S-ST group had a better score than the S-TS in the Initial Test (55% vs. 2%). However, the S-TS group showed better retention rates than the S-ST group both in Posttest 1 (84.2% vs. 53.2%), and in Posttest 2 (55% vs. 43.5%).

5 Discussion

The S-ST group showed better scores than the S-TS group in the Initial Test (55% vs. 2%). Though the S-TS group took only one study session before the Initial Test, the S-ST group took two study sessions before the Initial Test. Thus, the total amount of study time of the S-ST group (9 minutes and 20 seconds) doubled that of the S-TS group (4 minutes and 40 seconds). In addition, the S-TS group had taken the Initial Test a week after the first study session. On the contrary, the S-ST group took the restudy session soon before the Initial Test. Thus, a wide gap in scores between the S-TS and the S-ST group in the Initial Test was observed.

However, this results were reversed in Posttest 1 and Posttest 2 because the S-TS group had a significantly better score than the S-ST group both in Posttest 1 (84.2% vs. 53.2%) and in Posttest 2 (55% vs. 43.5%). As previous studies (Arnold & McDermott, 2013; Kornell, 2009) have found, the results of Posttest 1 and 2 revealed that the indirect effects of testing can be applied to L1–L2 pair learning.

The data suggested that there were indirect effects of testing for the items learners failed to recall. Indeed, 17 out of 23 participants in the S-TS group recalled nothing in the Initial Test. The average score was 0.82 out of 40 points (2%). Nevertheless, this result was reversed in Posttests 1 and 2. This suggested that learners in the S-TS group were able to benefit from the indirect effects of testing even though they had a poor performance in the Initial Test.

There are a number of reasons why the S-TS group was more successful in long-term retention of the meanings of the target items than the S-ST group. The participants in the S-TS group took the Initial Test, but they forgot most of the target words because there was a 1-week interval between the first learning session and the Initial Test. Therefore, the average scores were less than 1 point ($M = 0.82$ out of 40). They failed to retrieve the learned items from their memory in the Initial Test. However, the test may have prevented the learners in the S-TS group from being overconfident and helped them evaluate their learning achievement objectively. This should have led them to apply a great amount of effort in the restudy session. They were able to notice what words they had not recalled in the Initial Test, and were able to focus on these words during the restudy session. On the contrary, the participants in the S-ST group were not able to make the most of the indirect effects of testing because they did not have any opportunities to take the subsequent learning session after the Initial Test. They were not given a chance to make use of the information they had obtained from the test. This is likely to be the main reason why the S-TS group outperformed the S-ST group in Posttests 1 and 2.

Some limitations of this study should be noted. First, we did not measure in advance the participants' ability to remember new words. Thus, there might have been some differences in ability to remember new words between groups from the beginning. Second, we also did not measure their proficiency of the target language. It was unclear whether the participants truly had not known the target English words before the experiment. If some participants already knew some of the target words, this would affect the results.

To reinforce the finding of this study, a further study should examine what types of indirect effects of testing learners benefited from in the restudy session. A questionnaire survey is one of the best ways to deal with this issue. Some typical example of the questions are asking what participants thought when they got a lower score in the first test or what they focused on in the second learning session.

6 Conclusions

The main purpose of this study was to examine whether the indirect effects are useful even if learners experience lower score in a test (Initial Test). The answer is affirmative. This study revealed that the S-TS group had a better score than the S-ST group in Posttests 1 and 2, suggesting that learners can benefit from the indirect effects of testing even if they are not successful in the test. Such effects are powerful tools to establish a strong connection between L1 and L2 pairs. English teachers should be recommended to give their students the test before the restudy session. It was found that tests make a crucial contribution to long-term L2 vocabulary retention even if learners experience poor performance in the test, as long as a restudying session soon after a test is given.

The findings of this study are applicable to the actual English classroom. When teachers organize vocabulary quizzes and design their lessons, they should take the indirect effects of testing into consideration. After the test, the teachers should give their students adequate time to understand the test results, and should conduct an immediate restudy session. If this procedure is followed, the students are able to make better use of the indirect effects of testing even if their test scores are low.

References

- Agarwal, P.K., Finley, J.R., Rose, N.S., & Roediger, H.L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 25(6), 764–771. <https://doi.org/10.1080/09658211.2016.1220579>.
- Arnold, K.M., & McDermott, K.B. (2013). Test-potentiated learning: Distinguish between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 940–945. <https://doi.org/10.1037/a0029199>.
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57(1), 35–56. <https://doi.org/10.1111/j.1467-9922.2007.00398.x>.

- Blunt, J.R., & Karpicke, J.D. (2014). Learning with retrieval-based concept mapping. *Journal of Education Psychology, 106*(3), 849–858. <https://doi.org/10.1037/a0035934>.
- Brame, C.J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE-Life Sciences Education, 14*, 1–12. <https://doi.org/10.1187/cbe.14-11-0208>.
- Brown, P.C., Roediger, H.L., & McDaniel, M.A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.
- Butler, A.C. & Roediger, H.L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527. <https://doi.org/10.1080/09541440701326097>.
- Carpenter, S.K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin and Review, 13*(5), 826–830. <https://doi.org/10.3758/BF03194004>.
- Carpenter, S.K., Pashler, H., Wixted, J.T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory and Cognition, 36*(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition, 20*, 633–642. <https://doi.org/10.3758/BF03202713>.
- Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*, 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>.
- Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235. [https://doi.org/10.1002/\(SICI\)1099-0720\(200005/06\)14:3%3C215::AID-ACP640%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3%3C215::AID-ACP640%3E3.0.CO;2-1).
- Halamish, V., & Bjork, R.A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 801–812. <https://doi.org/10.1037/a0023219>.
- Hays, M.J., Kornell, N., & Bjork, R.A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290–296. <https://doi.org/10.1037/a0028468>.
- Kanayama, K., & Kasahara, K. (2015). The effect of word retrieval on L2 vocabulary learning. *Journal of the Hokkaido English Language Education Society, 15*, 21–33.
- Karpicke, J.D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486. <https://doi.org/10.1037/a0017341>.
- Karpicke, J.D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163. <https://doi.org/10.1177/0963721412443552>.

- Karpicke, J.D., & Grimaldi, P.J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24*, 401–418. <https://doi.org/10.1007/s10648-012-9202-2>.
- Karpicke, J.D., & Roediger, H.L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151–162. <https://doi.org/10.1016/j.jml.2006.09.004>.
- Kornell, N., Hays, M.J., & Bjork, R.A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998. <https://doi.org/10.1037/a0015729>.
- McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007a). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4/5), 494–513. <https://doi.org/10.1080/09541440701326154>.
- McDaniel, M.A., Roediger, H.L., & McDermott, K.B. (2007b). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review, 14*(2), 200–206. <https://doi.org/10.3758/BF03194052>.
- Meyer, A.N.D., & Logan, J.M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging, 28*(1), 142–147. <https://doi.org/10.1037/a0030890>.
- Nation, I.S.P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Putnam, A.L., Nestojko, J.F., & Roediger, H.L. (2016). Improving student learning. Two strategies to make it stick. In J.C. Horvath, J. Lodge, & J.A.C. Hattie (Eds.), *From the laboratory to the classroom: Translating the science of learning for teachers* (pp. 94–121). Oxford, UK: Routledge.
- Richland, L.E., Kornell, N., & Kao, L.S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243–257. <https://doi.org/10.1037/a0016496>.
- Roediger, H.L., Agarwal, P.K., McDaniel, M.A., & McDermott, K.B. (2011a). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382–395. <https://doi.org/10.1037/a0026252>.
- Roediger, H.L., & Karpicke, J.D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>.
- Roediger, H.L., & Karpicke, J.D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Roediger, H.L., & Marsh, E.J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1155–1159. <https://doi.org/10.1037/0278-7393.31.5.1155>.
- Roediger, H.L., Putnam, A.L., & Smith, M.A. (2011b). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 1–36). Oxford, UK: Elsevier.

- Soderstrom, N.C., & Bjork, R.A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, *73*, 99–115. <https://doi.org/10.1016/j.jml.2014.03.003>.
- Son, L.K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1980 to the present (and beyond). In J. Dunlosky & R.A. Bjork (Eds.), *A hand-book of memory and metamemory* (pp. 333–351). Hillsdale, NJ: Psychology Press.
- Tullis, J.G., Finley, J.R., & Benjamin, A.S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory and Cognition*, *41*, 429–442. <https://doi.org/10.3758/s13421-012-0274-5>.
- Vojdanoska, M., Cranney, J., & Newell, B.R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*, 1183–1195. <https://doi.org/10.1002/acp.1630>.

Vocabulary Learning Strategies and Vocabulary Size: Insights from Educational Level and Learner Styles

Alaa Alahmadi^a, Christopher Shank^a, and Anouschka Foltz^b

^a*School of Languages, Literatures and Linguistics, Bangor University, Bangor, Gwynedd, UK*

^b*Department of English Studies, University of Graz, Graz, Austria*
<https://doi.org/10.7820/vli.v07.1.alahmadi>

Abstract

This study investigates the effect of different vocabulary learning strategies (VLS) as well as different learner styles on vocabulary size in Saudi Arabic-speaking students in higher education. The goals of this study were to examine which VLS undergraduates used more frequently than postgraduates and vice versa, to determine which VLS related positively and significantly to vocabulary size, and to explore individual learner styles and their relationship to vocabulary size. Participants filled in a VLS questionnaire and completed a vocabulary size test. The results indicated that undergraduates tended to use simpler strategies than postgraduates. The strategies of *guessing a word's meaning from context* and *watching television* related positively with vocabulary size in both groups. Clustering analysis revealed two learner groups which differed in how frequently they used VLS overall, rather than in terms of which VLS they preferred. Those students who used more VLS overall also had larger vocabulary sizes, irrespective of educational level. We thus found no evidence for differences in individual learner styles in the current groups. We conclude that VLS usage should be encouraged overall, but that the need for instructors to cater to individual vocabulary learning styles may not be warranted.

Key words: Vocabulary acquisition; vocabulary learning strategies; vocabulary size; postgraduates; Arabic learners of English.

1 Background

In the past three or more decades, the importance of second language (L2) vocabulary learning has gained increased attention. Researchers, teachers and curriculum designers agree that acquiring the vocabulary of a foreign language is important for language learners (Coady, 1997; Gu, 2003; Nation, 1990; Ruutemets, 2005; Schmitt, 1997, 2000). Moreover, language learners seem to acknowledge the significance of vocabulary knowledge, as they usually use dictionaries rather than grammar books (Wilkins, 1972). In the same vein, Wilkins (1972) stated that “without grammar very little can be conveyed, without vocabulary nothing can be conveyed” (p. 110–111). Furthermore, McCarthy (1990) and Shen (2008) argued that language proficiency is heavily

dependent on individuals' vocabulary knowledge. Indeed, several studies have found a significant and positive correlation between learners' vocabulary size and scores on formal tests of the four language skills: listening, speaking, reading and writing (Koizumi & In'nami, 2013; Laufer, 1994, 1997; Meara & Jones, 1988; Stæhr, 2008, 2009).

While vocabulary knowledge plays a vital role in language proficiency, many researchers have argued that vocabulary acquisition is the most challenging feature of learning a foreign language (Milton, 2009; Schmitt, 2000). McCarthy (2001, p. 2) suggests that "vocabulary forms the biggest part of the meaning of any language, and vocabulary is the biggest problem for most learners". To address these difficulties, a number of researchers have developed various strategies to support language learners in their efforts to efficiently acquire vocabulary (Gu, 2003; Oxford, 1990; Read, 1997; Schmitt, 1997).

1.1 Language Learning Strategies and Vocabulary Learning Strategies

Language learning strategies (LLS) have received considerable attention since the 1960s (Safian, Malakar, & Kalajahi, 2014). This reflects the educational shift from focusing on instructors and teaching methods to learners and learning styles (Chamot & O'Malley, 1987; Fillmore, 1983; Stern, 1975; Wenden, 1982). Schmitt (1997) attributed this educational shift to the general awareness that aptitude is not the only major factor for successful language learning, but that individual learner's strategies may be equally important. As such, researchers have shifted their focus to learners' individual learning approaches and how learners control their learning and language use.

There are various, and sometimes controversial, definitions of LLS in the literature (Oxford, 1990). We follow Cohen's (1998) definition of LLS:

Language learning and language use strategies can be defined as those processes which are consciously selected by learners and which may result in action taken to enhance the learning or use of a second or foreign language, through the storage, retention, recall, and application of information about that language. (p. 4)

While VLS are considered a sub-class of LLS, O'Malley and Chamot (1990) have argued that most LLS are used for accomplishing vocabulary learning tasks. We follow Catalan's (2003) definition of VLS as:

the mechanism used in order to learn vocabulary as well as steps or actions taken by students (a) to find out the meaning of unknown words, (b) to retain them in long-term memory, (c) to recall them at will, and (d) to use them in oral or written mode. (p. 56)

Many vocabulary learning strategies have been proposed in the literature (e.g., Nation, 2001; Schmitt, 1997; Stoffer, 1995). This study uses Schmitt's (1997) comprehensive and frequently cited VLS taxonomy, which integrates key components of Nation's (1990), Oxford's (1990) and Cook and Mayer's (1983) taxonomies. Schmitt's taxonomy (see Figure 1) divides VLS into discovery strategies,

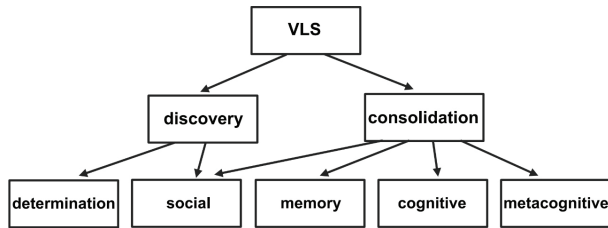


Figure 1. Graphical Overview of Schmitt's (1997) Taxonomy

which are used to learn the meaning of a new word, and consolidation strategies, which are used to remember words once they have been initially learnt. Schmitt's discovery strategies have two sub-strategies: determination strategies and social strategies. Determination strategies aid vocabulary acquisition by providing a set of limited choices from which a word's meaning can be determined, such as using a dictionary or deriving meaning from context. Social strategies support vocabulary attainment by cooperating with others in the acquisition process (Tanyer & Ozturk, 2014). Consolidation strategies are further subdivided into social (see above), memory, cognitive and metacognitive strategies. Memory strategies refer to learners' mental attempts to link new words with their background knowledge (Schmitt, 1997). Cognitive strategies involve manipulating language materials to enhance the learning process, for example, note-taking, analysis or outlining (Oxford, 2003). Finally, metacognitive strategies may be described as knowledge about learning. This could involve learners' conscious knowledge about how to identify their own learning style preferences, monitoring learning shortcomings and evaluating their learning progress (Oxford, 2003; Schmitt, 1997).

1.2 The Importance of Vocabulary Knowledge

Previous studies have shown a statistically significant positive relationship between vocabulary knowledge and language proficiency (e.g., Milton, 2009). Most research concerning this relationship has been conducted within the realm of reading (Stæhr, 2008). A number of researchers have proposed vocabulary sizes necessary to achieve a minimum level of understanding in a foreign language. Milton (2009) indicated that English as a Foreign Language (EFL) learners needed to know at least 3000 words to perform well in basic communicative tasks (Nation, 1990). Proposed thresholds for reading academic and authentic texts range from 5000 words to 9000 word-families (headword plus inflected forms and some derived forms; Al-Masrai & Milton, 2012; Hirsh & Nation, 1992; Nation, 2001). Nation (2006) suggested a vocabulary knowledge of 6000 to 7000 word-families to understand spoken discourse. Overall, it is presumed that, regardless of the skill, learners familiar with more words have more opportunities to achieve better comprehension and production of the foreign language.

A number of studies have investigated how English learners' VLS use relates to vocabulary size, using VLS questionnaires in combination with tests that estimate vocabulary knowledge. These studies have yielded rather diverse results. For example, Al Qahtani (2005) conducted a comprehensive analysis of VLS use

among 455 students from three educational levels (ages 13 through 15, high school and university undergraduate level) in Saudi Arabia and found significant correlations between vocabulary size and guessing the meaning of a word from its structure, monolingual dictionary usage and learners' self-monitoring by listening to their own recordings to detect errors.

Hamzah, Kafipour and Abdullah (2009) have also investigated VLS usage and its relationship with vocabulary size in 125 Iranian second-year undergraduates majoring in Teaching English as a Foreign Language (TEFL). In their study, significant and positive correlations were found between vocabulary knowledge and performing a physical action while learning a new word, communicating with native speakers and watching English media.

Alsaif (2011) explored the use of VLS and their relationship with vocabulary size among 111 Saudi male students representing four public schools and distinctive school levels (from level 7 to level 11). The findings highlighted a positive relationship between vocabulary knowledge and associating new words with known synonyms and antonyms and guessing the words' meanings from context.

Finally, Tanyer & Ozturk (2014) examined 80 Turkish university students in years 1 through 4 of their undergraduate studies, who were majoring in English Language Teaching (ELT). A hierarchical multiple regression analysis revealed that social, cognitive and metacognitive strategies significantly influenced participants' vocabulary knowledge and explained 6.5%, 5.3% and 5.2%, respectively, of the variability in vocabulary size.

1.3 Individual Learner Styles

Individual learning styles refer to the idea that learners differ in terms of the approach of teaching or learning that is optimal for them (Pashler, McDaniel, Rohrer & Bjork, 2008). The concept of learning styles has gained much attention recently (e.g., Kozhevnikov, 2007; Sternberg, Grigorenko, & Zhang, 2008), which has led learning styles proponents to design learning styles models or schemes, such as Kolb's (1984, 1985) Learning Styles Inventory. These schemes can assist instructors in highlighting their students' learning styles. Based on this, teachers can also adapt their instruction to these individual learner styles. For example, Kolb's (1985) learning styles differentiate between a preference for active experimentation and reflective observation. In terms of VLS, this would reflect a preference for strategies that involve active participation compared to observation or for strategies that involve speaking compared to listening.

However, whether or not using individual learning styles as a means to improve learning outcomes is effective has been debated in the literature. Ormrod (2008) suggests that "some cognitive styles and dispositions do seem to influence how and what students learn" (p. 160). Kojic-Sabo and Lightbown (1999) explored how English as a Second Language (ESL) and EFL learners' styles relate to vocabulary knowledge. They used clustering analysis to group participants in terms of learner styles and found eight different learner groups. Their results suggest that learners who make more use of VLS have larger vocabulary knowledge. In addition, the two most successful groups in terms of vocabulary knowledge

were characterised by a large amount of time spent on using strategies and a high amount of strategy use outside of the classroom. Similarly, Pashler et al.'s (2008) review of learning styles suggests that there was no adequate evidence to justify the prominent role of individual learning styles in general educational practice.

1.4 Current Study

The purpose of this study is to shed light on the VLS usage of students in higher education. We will consider the VLS usage of undergraduates who are completing their degree in Saudi Arabia and postgraduates studying in an L2 environment, as this is a common educational path for Saudi students. Universities in English-speaking countries are accepting many international postgraduates into their programmes. These students are typically required to have reached a certain level of English proficiency, typically a score of 5.5 or above in the International English Language Testing System (IELTS), before they can begin their studies. In addition, these students need to build up their vocabulary competence in order to participate effectively within their academic disciplines in the L2. Postgraduates in the L2 environment have different opportunities to practice their vocabulary than students in the first language (L1) environment. In particular, they have more opportunities to learn vocabulary through social strategies, such as communicating with and listening to native speakers. This study investigates which VLS postgraduates use significantly more frequently than undergraduates and vice versa and which VLS are related to participants' vocabulary size. Furthermore, the study explores whether we can identify any individual learner styles and how different learner styles relate to vocabulary size. The latter analysis follows Kojic-Sabo and Lightbown (1999) in using cluster analysis to explore whether there are any distinct VLS profiles that could contribute to participants' vocabulary knowledge. To the best of our knowledge, very few previous studies have attempted to explore individual learner styles in terms of L2 vocabulary acquisition, and this is the first study that considers VLS usage of postgraduate learners. In summary, the study attempts to answer the following research questions (RQs):

1. Which of the VLS relate significantly to vocabulary size in both groups?
2. Are there any strategies that postgraduates use significantly more or less frequently than undergraduates?
3. Are there distinct VLS profiles or learning styles across the sample and, if so, which of these contribute to vocabulary size?

2 Study 1

Study 1 explores VLS and the breadth of vocabulary knowledge in undergraduates studying in their home country.

2.1 Methodology

2.1.1 Participants

A total of 49 students (mean age = 23.65, standard deviation [*SD*] = 2.445) participated in the study. All students were third year (level six) students in the

English Department at King Abdulaziz University in Saudi Arabia. Students received approximately 1600 h of EFL instruction during their public school and university education (Alqurashi, 2013).

2.1.2 Materials

VLS questionnaire. The questionnaire we used to gauge students' VLS use was adapted from Alsaif (2011) and based on Schmitt's VLS taxonomy (1997). It contained 37 closed questions. An additional open question that allowed participants to mention additional strategies that they used will not be reported here because no learner mentioned additional learning strategies. All closed questions used the following Likert scale to gauge how often learners used a particular learning strategy: always = 4, often = 3, sometimes = 2, rarely = 1 and never = 0.

Following Alsaif (2011), the questionnaire was organised into three sections: (1) strategies used to learn new words, (2) strategies used to consolidate already learned words and (3) general VLS. Each of the learning strategies was categorised as either memory, social, cognitive, metacognitive or determination strategies following Schmitt (1997).

Vocabulary size test. We used Meara and Milton's (2003) vocabulary size test X_Lex (paper-based format; see Milton, 2009) to gauge participants' vocabulary size. This test estimates participants' passive vocabulary knowledge in terms of breadth up to a vocabulary knowledge of 5000 words and has a high level of test validity and reliability (Al-Mutawa, 2013). We used this test as participants' vocabulary knowledge was expected to not exceed 5000 words (e.g., Al-Akloby, 2001; Al-Hazemi, 1993; Alsaif, 2011).

X_Lex is a checklist vocabulary test in which participants place a checkmark next to the words they know. The test consists of six columns with 20 words each (120 words in total), representing the 5000 most frequent words in English. The words were selected from Hindmarsh's (1980) and Nation's (1984) frequency lists (Milton, 2009). The first five columns contain real words that are among the most frequent 5000 words in English. The last column contains non-words or pseudo-words to gauge the amount of guessing by the learners. Vocabulary size scores are calculated as follows: All checkmarks within the first five columns (which contain real words) are added up and this number is multiplied by 50. Then, all checkmarks in the sixth column (which contains pseudo-words) are added up and multiplied by 250. Finally, the total number for the sixth column is subtracted from the total number for the first five columns. For example, if the first five columns yield a score of 2500 and the sixth column yields a score of 1000, then the participants' vocabulary size score is 1500.

2.1.3 Procedure

Participants first completed the VLS questionnaire, which was administered in Arabic, the participants' L1. After a short break, participants completed the vocabulary size test. There was no time limit for completion of the questionnaire and vocabulary size test, but all the participants finished within 20 min.

2.2 Results

2.2.1 Summary Measures

The results of the X_Lex vocabulary size test showed that participants' estimated mean vocabulary size is 1976 words ($SD = 597$). Estimated vocabulary size ranged from 350 words to 2900 words out of a possible maximum test score of 5000 words. The ratings for individual VLS range from 0.85 (a value between *rarely used* and *never used* on the scale) to 2.95 (a value corresponding to *often*), with a mean rating of 1.98 (a value corresponding to *sometimes*) out of a maximum score of 4.00.

2.2.2 Relationship between VLS and Vocabulary Size

To determine which VLS contribute significantly to vocabulary size (RQ1), we performed multiple regression analyses, separately for each type of VLS in Schmitt's taxonomy (determination, cognitive, social, memory and metacognitive strategies). All analyses had vocabulary size as the dependent variable and all VLS of the relevant type as independent variables. All independent variables were centred before analysis to minimise collinearity (Belsley, Kuh, & Welsh, 2005). Independent variables that did not contribute significantly to model fit were removed in a stepwise procedure to yield the final analysis model (Baayen, 2008). Table 1 shows the results from the final statistical models. The only strategy that is significant at the level of $p < 0.001$ is *guessing the meaning of new words from context*. Overall, we can see that most of the strategies that are related to vocabulary size were categorised by Alsaif (2011) as strategies to learn new words (7, 9, 10, 14, 19, and 20) rather than as strategies to consolidate and memorise learned words (21 and 24) or general VLS (32). In addition, the most frequent strategy type in Table 1 is memory strategies (10, 14, 19, and 20).

2.3 Discussion

The results revealed that undergraduate participants' receptive vocabulary knowledge is generally poor and would likely not allow learners to perform well in basic communicative tasks and general reading tasks. In addition, participants did not employ many vocabulary learning strategies on a regular basis. However, three of the VLS that positively relate to students' vocabulary size had mean ratings above 2 (i.e., above *sometimes*) and were thus among those that undergraduates used more frequently. This suggests that some strategies that students were using might have contributed to building their vocabulary knowledge, possibly because use of these strategies might indicate their engagement with English in general. Those findings will be discussed in more detail in the general discussion section.

3 Study 2

Study 2 tests VLS and the breadth of vocabulary knowledge in postgraduates living in an L2 environment.

Table 1. Results from the Multiple Regression Analyses for Undergraduates

Strategy type	VLS	Mean (SD)	Estimate	Std. error	t-value	p-value
Determination	9. Guessing the meaning of the new words from their contexts	2.35 (1.332)	235.75	51.86	4.545	< 0.001***
Cognitive	21. Writing the words in a special vocabulary notebook	2.27 (1.204)	155.28	60.39	2.571	< 0.05*
Social	7. Asking the teacher about vocabulary meanings	1.24 (1.267)	183.23	80.55	2.275	< 0.05*
Memory	10. Using grammar cues to guess the meaning of words	2.27 (1.335)	199.36	57.97	3.439	< 0.01**
	14. Writing the words in full sentences	1.92 (1.367)	156.51	52.53	2.979	< 0.01**
	19. Associating new words with known synonyms	0.86 (1.225)	212.23	68.03	3.120	< 0.01**
	20. Associating the new words with known opposites	2.27 (1.204)	-178.32	72.61	-2.456	< 0.05*
	24. Writing the words with all the synonyms known	1.63 (1.185)	112.65	64.07	1.758	= 0.09
Metacognitive	32. Watching television programmes	1.53 (0.739)	152.66	57.70	2.646	< 0.05*

3.1 Methodology

3.1.1 Participants

The postgraduate group included 22 male Saudi learners (mean age = 30.50, $SD = 3.051$) completing their Master's or PhD degrees in different disciplines at Bangor University. These participants have spent between 2 to 7 years in an L2 environment.

3.1.2 Materials

VLS questionnaire. We used the same questionnaire as in Study 1 to measure the participants' VLS.

Vocabulary size test. Participants' passive vocabulary size was measured using the XK_Lex vocabulary size test developed by Al-Masrai and Milton (2012; see Al-Masrai, 2009, for validity and reliability information), which captures vocabulary knowledge beyond the 5000-word limit and is thus appropriate for learners in

the L2 environment at a higher educational level. Similar to the X_Lex test used in Study 1, the XK_Lex is a paper-and-pencil checklist vocabulary test in which participants place a checkmark next to the words they know. It comprises 100 words representing the 10 000 most frequent words of English, divided into 10 columns with 10 words each. The lexical items in the test are taken from Nation (1984) and Kilgarriff (2006). To minimise the effect of guessing, each column of the test also includes two pseudo-words, for a total of 20 pseudo-words. Similar to the X_Lex test in Study 1, vocabulary size scores are calculated by adding up all the real words that received a checkmark and multiplying the sum by 100. Then, all checked pseudo-words are added up and the sum is multiplied by 500. Participants' vocabulary size score was derived by subtracting the latter sum from the first.

3.1.3 Procedure

The procedure was the same as in Study 1.

3.2 Results

3.2.1 Summary Measures

The XK_Lex test revealed that postgraduates' estimated mean vocabulary size is 5368 words ($SD = 1307$). Vocabulary sizes ranged from 3100 to 8200 words out of a possible maximum test score of 10 000 words. Mean VLS ratings ranged from 1.09 (roughly corresponding to *rarely* on the scale) to 3.41 (a value that is between *often* and *always* on the scale), with an overall mean rating of 2.11 (which roughly corresponds to *sometimes*) out of a maximum rating of 4.00.

3.2.2 Relationship between VLS and Lexical Knowledge

We performed multiple regression analyses analogous to the ones in Study 1 to determine which VLS relate significantly to vocabulary size (RQ1) in Saudi postgraduates. Table 2 presents the results from the final statistical models. Again, most VLS strategies that relate to vocabulary size are strategies used to acquire new words (2, 9, 11, 12, 13, 19 and 20) rather than strategies to consolidate and memorise learned words (none) or general VLS (32, 34 and 37). Furthermore, the most common strategy category in Table 2 is again memory strategies (11, 12, 13, 19 and 20). However, three of the five memory strategies which relate to vocabulary size show a negative relationship, such that more use of the strategy relates to a smaller vocabulary. All three of these strategies relate to considering words in isolation rather than in context.

3.2.3 Comparison of VLS Use between Groups

In this section, we will highlight VLS that postgraduates use significantly more or less frequently than undergraduates (RQ2). Table 3 lists all VLS for which the mean ratings for undergraduates were significantly or marginally higher than the

Table 2. Results from the Multiple Regression Analyses for Postgraduates

Strategy type	VLS	Mean (SD)	Estimate	Std. error	t-value	p-value
Determination	9. Guessing the meaning of the new words from their contexts	2.64 (0.790)	883.3	343.1	2.574	< 0.05*
Cognitive	2. Reading the text aloud before searching for new words meanings	1.41 (0.854)	628.1	261.6	2.401	< 0.05*
Social	37. Speaking English with non-Arabic speakers in shops, hospitals, restaurants, etc.	3.32 (1.041)	510.4	256.5	1.989	= 0.06
Memory	11. Visualising the meaning of the words	1.86 (0.941)	869.2	262.0	3.317	< 0.01**
	12. Saying the words repeatedly to learn their sounds	2.59 (1.182)	-422.8	194.6	-2.173	< 0.05*
	13. Writing the words alone repeatedly to learn their written forms	2.14 (0.889)	-783.1	235.3	-3.328	< 0.01**
	19. Associating new words with known synonyms	2.23 (1.27)	-440.7	205.7	-2.143	< 0.05*
	20. Associating the new words with known opposites	2.23 (1.11)	752.5	255.6	2.944	< 0.01**
Metacognitive	32. Watching television programmes	3.44 (0.590)	650.3	218.2	2.981	< 0.01**
	34. Reading newspapers or magazines	2.41 (0.908)	601.8	261.9	2.298	< 0.05*

Table 3. Vocabulary Learning Strategies for Which Mean Ratings for Undergraduates Were Significantly or Marginally Higher than Mean Ratings for Postgraduates

VLS	Mean scores (SD) undergraduates	Mean scores (SD) postgraduates	t-test
7. Asking the teacher about their meanings	2.86 (0.979)	1.59 (0.959)	t = -5.1113 p < 0.001***
16. Volunteering to say the words loudly in class if the teacher asks	2.24 (1.217)	1.50 (1.225)	t = -2.3747 p = 0.09
25. Writing the words with all the opposites I know	1.67 (1.144)	1.09 (0.921)	t = -2.2805 p = 0.1

mean ratings for postgraduates. Two-tailed *t*-tests with *p*-values adjusted for multiple comparisons with a false discovery rate correction (Benjamini & Hochberg, 1995) showed that only one strategy was used significantly, and a further two marginally more frequently, by undergraduates compared to postgraduates, reflecting the infrequent use of VLS in general among undergraduate participants. Notably, two of the strategies that undergraduates used more frequently than postgraduates are social strategies involving simple classroom interaction (7 and 16). All strategies were relatively simple and focused on the word in isolation rather than in context.

In contrast, Table 4 lists all VLS for which the mean ratings for postgraduates were significantly or marginally higher than the mean ratings for undergraduates. Two-tailed *t*-tests with *p*-values adjusted using a false discovery correction showed that seven strategies were used significantly or marginally more frequently by postgraduates compared to undergraduates. Notably, all but two of these strategies serve to consolidate and memorise already learned words. Moreover, postgraduates use strategies that allow them to process words in their sentential and conversational contexts (2, 34, 35, 36 and 37) more frequently than undergraduates. Finally, some of the strategies that postgraduates use more frequently than undergraduates are related to living in an L2 environment and the opportunities to engage with the L2 that go along with this (36, 37 and possibly 34, 35).

3.2.4 Individual Differences

This section focuses on individual learner styles (RQ3) using cluster analysis. This procedure allows us to group participants with similar VLS profiles and

Table 4. Vocabulary Learning Strategies for which Mean Ratings for Postgraduates were Significantly or Marginally Higher than Mean Ratings for Undergraduates

VLS	Mean scores (SD) undergraduates	Mean scores (SD) postgraduates	<i>t</i> -test
2. Reading the whole text aloud before searching for the meaning of the new words	1.57 (1.323)	2.27 (0.985)	$t = 2.4827$ $p = 0.08$
6. Using English/English dictionary	1.24 (1.269)	2.41 (1.221)	$t = 3.6714$ $p < 0.01^{**}$
31. Connecting the English words with other words in English which have the same sound, like <i>meat</i> and <i>meet</i> , etc.	1.53 (0.737)	2.59 (1.141)	$t = 3.9998$ $p < 0.01^{**}$
34. Reading newspapers or magazines	1.55 (1.276)	2.41 (0.908)	$t = 3.2271$ $p < 0.05^*$
35. Reading stories or novels	1.41 (1.257)	2.14 (1.167)	$t = 2.3735$ $p = 0.09$
36. Speaking English with my friends inside or outside school	1.57 (1.258)	2.64 (1.002)	$t = 3.814$ $p < 0.01^{**}$
37. Speaking English with non-Arabic speakers in shops, hospitals, restaurants, etc.	2.53 (1.386)	3.32 (1.041)	$t = 2.6475$ $p = 0.07$

has the ability to uncover different learner styles. We used a K-means clustering approach, an unsupervised machine learning algorithm that identifies a predetermined number of groups in the data. One issue with this approach is to determine the right number of clusters (K) for the analysis. We did this using the NbClust package in R (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). The NbClust function uses 30 different approaches for finding the optimal number of clusters in a data set and determines which number of clusters is considered optimal in the largest number of approaches. The optimal number of clusters was 2. Inspection of the groups suggests that the first cluster represents learners with infrequent VLS usage overall (24 undergraduates and 10 postgraduates). The second cluster includes learners with frequent VLS usage overall (25 undergraduates and 12 postgraduates). In particular, the mean usage ratings for all VLS were numerically lower and the mean usage ratings for 26 of the 37 VLS were statistically significantly lower (Welch two-sample t -tests with p -values adjusted with a false discovery rate correction, all $t > 2$, all $p < 0.05$) for participants in the first cluster than for participants in the second cluster.

Next, a linear regression model tested whether the level of education and the VLS profile (as established through the cluster analysis) relate to vocabulary size. The model included level of education (undergraduate vs. postgraduate), the VLS profile (low vs. high VLS use) and their interaction as independent variables and vocabulary size as dependent variable. The interaction did not significantly contribute to model fit and was removed. The results show a significant main effect for both level of education (estimate = -3375.2, std. error = 213.7, $t = -15.478$, $p < 0.001^{***}$) and VLS profile (estimate = -496.9, std. error = 197.8, $t = -2.102$, $p < 0.05^*$). This suggests that using more VLS in general is related to larger vocabulary size independently of level of education.

3.3 Discussion

Results from Study 2 indicated that Saudi postgraduates' vocabulary size is considerably higher than that of undergraduates. Postgraduate vocabulary size showed significant relationships with several strategies that involve considering words in their sentential and conversational context. In addition, VLS used more frequently by undergraduates tended to focus on words in isolation and simple classroom interactions, whereas VLS used more frequently by postgraduates tended to focus on words in their sentential and discourse contexts. Finally, we determined two groups of learners which differed in their overall use of VLS rather than in preferring certain VLS over others. These findings are discussed in more detail in the next section.

4 General Discussion

4.1 Participants' Vocabulary Size

The findings of this study indicate that undergraduates' and postgraduates' vocabulary sizes are comparable to findings from previous studies that measured Saudi students' vocabulary size (e.g., Al-Akloby, 2001; Al-Hazemi, 1993;

Al Qahtani, 2005; Alsaif, 2011; Masrai, 2015). None of the undergraduates met the suggested vocabulary thresholds for performing well in basic communication, general reading comprehension, reading academic or authentic texts, comprehending authentic materials or understanding spoken discourse (Al-Masrai & Milton, 2012; Hirsh & Nation, 1992; Laufer, 1997; Milton, 2009; Nation, 2006). All postgraduates met the threshold for general reading comprehension and performing well in basic communication (Laufer, 1997; Milton, 2009). In addition, a majority of postgraduates met Hirsh and Nation's (1992) suggested thresholds for reading academic and authentic texts, reflecting that their vocabulary knowledge is adequate. Our results support Fu's (2005) argument that lexical attainment is often problematic for many L2 students, even for advanced learners.

Previous studies have linked the low vocabulary size among Saudi students to the late beginning of English instruction within Saudi public schools (grade 4, 10 years old), the classroom environment, instructional approaches, teachers' experience, word difficulty variables and morphological processing (Alsaif, 2011; Masrai, 2016; Masrai & Milton, 2015). The results from the current studies suggest that the infrequent use of VLS may relate to Saudi undergraduates' low vocabulary uptake. The results also suggest that undergraduates may benefit from quite simple and easily implemented VLS, such as asking about a word's meaning in class. However, students may not always be aware of the benefits of such simple strategies. EFL instructors could therefore emphasise the importance of VLS to students or provide students with a repertoire of VLS that are relevant for their level of knowledge and that students can draw from when learning vocabulary.

4.2 Relationship between VLS Use and Lexical Knowledge

Our first research question (RQ1) explored which VLS relate positively and significantly to vocabulary size in both groups. These strategies are *guessing the meaning of words from context*, and *watching television programmes*. Notably, the strategy of *guessing the meaning from context* related to vocabulary size in both groups and was the only VLS that was significant at the $p < 0.001$ level in the undergraduate group. This finding supports previous studies, which also found significant relationships of this VLS and vocabulary size (e.g., Al Qahtani, 2005; Alsaif, 2011). Since guessing the meaning of words from context relates to vocabulary size in several studies, it is worth exploring whether using this strategy increases one's vocabulary size or whether participants with larger vocabulary sizes simply have more opportunities to use this strategy. In order to guess the meaning of words from context, learners need to have adequate knowledge about the vocabulary and grammatical structures in the context. Furthermore, the more words, grammatical structures and sentences in context learners understand, the easier it becomes for them to guess an individual unknown word from the context. Lexical inferencing strategies are also useful as test-taking strategies when using a dictionary is not an option. Thus, students with substantial English test-taking experience may be more effective users of this particular strategy. Even when dictionary use is an option, being able to guess the meaning of words from context, although possibly more error prone, is more time-efficient than looking up words in a dictionary. If this is the case, then learners with larger vocabularies

and more confidence to guess correctly may simply choose to employ this more time-efficient strategy more often than learners with smaller vocabularies. Alternatively, using this strategy may actually benefit vocabulary acquisition. In particular, in order to successfully guess the meaning of a word from context, the learner needs to engage with the material in depth (Ellis, 1995; Hulstijn, 2001). This involves engagement such as determining the meaning of surrounding words, the grammatical structures and thematic roles of the surrounding sentences, the topic and broader context of the text. Such deep engagement with the text may actively increase learners' vocabulary knowledge, possibly by strengthening associative bonds between words (Richards, 1976). However, previous studies on this topic have found inconclusive results: While Shangarfam, Ghorbani, Safarpour and Maha (2013) found an advantage for guessing word meaning over looking words up, Mondria (2003) did not.

Interestingly, another VLS which positively related to vocabulary knowledge in both groups is watching television programmes. This strategy actually requires learners to guess unknown words from context, as speech in television programmes is typically too fast to allow looking up vocabulary. In addition, it is possible that only students with sufficient confidence in their lexical inferencing ability attempt to watch English-language television. Overall, our results support a connection between guessing strategies and vocabulary size; however, more studies with larger sample sizes, especially at the postgraduate level, are needed to determine whether these strategies increase learners' vocabulary knowledge.

It is also noteworthy that three of the memory strategies in the postgraduate group related negatively to vocabulary size. All of these strategies involved basic strategies that relate to words in isolation rather than in context, such as *saying words repeatedly to learn their sounds* or *writing words repeatedly to learn their written forms*. It is very unlikely that using such strategies would decrease learners' vocabulary size. Rather it seems that learners with smaller vocabulary sizes to begin with may select these strategies more frequently than learners who already have larger vocabularies.

Our results also suggest that the interrelationship between vocabulary acquisition and VLS use might be more sophisticated than what is reported in the literature. In particular, only two strategies showed reliable positive relationships with vocabulary size in both groups of students. This suggests that there are some strategies which may be beneficial for certain learner groups or learners of certain proficiency levels, but not for others. In line with Alsaif (2011) and Alqurashi (2013), the most frequent type of strategy that related to learners' vocabulary size in the current study was memory strategies. However, not all memory strategies related positively to vocabulary size, with some of the more basic memory strategies associated with learners with small vocabularies. Contrary to results from Hamzah et al. (2009) and Tanyer & Ozturk (2014), only one social strategy in each group related to learners' vocabulary size. This suggests that the undergraduates, and even some postgraduates, in the current study focused on and may have benefited from strategies that encourage rote learning and memorisation. Further studies with larger sample sizes are needed to determine which strategies may potentially be beneficial for which learner groups.

4.3 Comparison of VLS Use across Groups

RQ2 focused on strategies which postgraduates use significantly more or less frequently than undergraduates. The current results revealed that undergraduates only used three strategies significantly or marginally more often than postgraduates. Moreover, all of the VLS are simple strategies that consider words in isolation and represent strategies that are typically part of the normal routine in the EFL classroom and thus represent expected classroom behaviour, such as volunteering to say words aloud in class or asking the teacher about a word's meaning. In contrast, postgraduates take advanced content classes in their field of study with little opportunity to volunteer to say words aloud in the class. In addition, students may consider asking the teacher about a word's meaning inappropriate in advanced content-based classes. This highlights that the learning environment in the classroom may be more or less conducive to the use of certain VLS. In addition, there is no evidence that undergraduates used any VLS that are more complex or occur outside of the classroom more frequently than postgraduates.

Postgraduates used seven strategies reliably or marginally more frequently than undergraduates. Again, some of these strategies seem to be related to postgraduates' current learning environment. As students in an L2 environment, postgraduates have ample opportunity to interact with native English speakers and other international students in English. Postgraduates also make use of advanced metacognitive and social strategies, such as reading newspapers or magazines and stories or novels or interacting with speakers in English inside and outside of the classroom. This suggests that postgraduates are actively engaging with authentic English language materials and English language speakers, which are more readily available in the L2 environment compared to the L1 environment of the undergraduates. In addition, postgraduates are more likely to use VLS outside of the classroom environment.

4.4 Individual Differences

We used cluster analysis to explore whether there were any distinct VLS profiles or learning styles (RQ3). Cluster analysis determined two VLS usage profiles: overall infrequent VLS usage and overall frequent VLS usage. We thus found no evidence for different types of learner styles, such as visual learners compared to auditory learners, a preference for speaking compared to listening, or a preference for social compared to memory strategies, when it comes to VLS. Our results thus do not support the claim that individual learner profiles "might do more justice to the individuality of the language learner" (Skehan, 1986, p. 82). These results also suggest that asking instructors to cater to individual vocabulary learner styles may not be warranted, an argument which is in line with that of Pashler et al. (2008). However, it may also be the case that the learners do have individual learner styles, but do not have sufficient awareness of their needs as a learner to select VLS that work well for their particular learner style. We did, however, find that participants who reported using VLS more frequently overall had significantly larger vocabulary sizes and that this effect occurred in addition to

whether they were undergraduates or postgraduates. This finding is in line with that of Sanaoui (1992) and Kojic-Sabo and Lightbown (1999), who concluded that frequent and elaborate usage of VLS was related to high achievement levels. Gu & Johnson (1996) also contended that both learners' vocabulary knowledge and language proficiency appear to be related to certain learning aspects, such as learners' motivation in language acquisition, the ability to use dictionaries and the willingness to spend additional time on practicing novel acquired lexical items. However, our findings contrast with that of Lessard-Clouston (1996), who did not find any relationship between frequent usage of VLS and language learning success. It is thus possible that both the quantity and quality of VLS are involved in language learning success, and that quantity alone may not necessarily relate to higher success in language learning.

Overall, the results of our study suggest that VLS usage in general positively relates to vocabulary size. It seems that learners who show greater engagement with VLS overall also have larger vocabularies, irrespective of educational level. Again, the directionality of this result is not clear. It is possible that engagement with the learning process through the frequent use of VLS might lead to larger vocabulary sizes (e.g., Hamzah et al., 2009; Tanyer & Ozturk, 2014). However, another possibility is that learners who already have larger vocabularies have more means to engage with VLS, especially those that require a certain vocabulary base, such as reading newspapers, interacting with other speakers outside of the classroom, etc.

5 Conclusion

This study highlighted the VLS use of students in higher education and the relationship of VLS with students' breadth of vocabulary knowledge. The results showed that postgraduates had overall larger vocabulary sizes and used more strategies that considered lexical items in their sentence and discourse contexts than undergraduates. The strategies of *guessing the meaning of words from context* and *watching television programmes* stood out among the VLS because they related positively to vocabulary size in both groups. Cluster analysis provided two VLS usage profiles which differed in the frequency of VLS use overall rather than in terms of individual learner styles. Frequency of VLS use and educational level were found to be independently related to learners' vocabulary size. The results tentatively suggest that VLS use overall should be encouraged in the EFL classroom, that guessing strategies may possibly contribute to vocabulary size in both undergraduates and postgraduates, and that students may not differ enough in terms of their learning styles to justify asking teachers to cater to individual learning styles.

Acknowledgements

The first author (Alaa Alahmadi) is immensely grateful to Dr. Ahmed Al-Masrai, Mr. Osama Alruwaili and Mr. Mohammed Hakami for their assistance and support during data collection.

References

- Al-Akloby, S. (2001). *Teaching and learning English vocabulary in Saudi Arabian public schools: An exploratory study of some possible reasons behind students' failure to learn English vocabulary*. Unpublished PhD thesis, University of Essex.
- Al-Hazemi, H. (1993). *Low level EFL vocabulary tests for Arabic speakers*. Unpublished PhD thesis, University of Wales, Swansea.
- Al-Masrai, A. (2009). *Measuring the English vocabulary size of Saudi university students: Validating a new 10,000 word vocabulary size test*. Unpublished Master dissertation, Swansea University, Swansea, UK.
- Al-Masrai, A., & Milton, J. (2012). The vocabulary knowledge of University Students in Saudi Arabia. *Perspectives (TESOL Arabia)*, 19(3), 13–19.
- Al-Mutawa, Y. (2013). *Estimating the EFL vocabulary size of learners in school and university in Saudi Arabia: Validity of X_Lex vocabulary size test*. Unpublished MA thesis, Swansea University.
- Al Qahtani, M. (2005). *The use of vocabulary learning strategies by EFL learners at three different educational levels*. Unpublished PhD thesis, Essex University.
- Alqurashi, N. (2013). *Vocabulary learning strategies of Saudi undergraduate EFL students at Taif University and its relation to their vocabulary size*. Unpublished MA dissertation, Anglia Ruskin University, Cambridge, UK.
- Alsaif, A. (2011). *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia*. Unpublished PhD thesis, Swansea University, Swansea, UK.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity (Vol. 571)*. Hoboken, New Jersey: John Wiley & Sons.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Catalan, R. (2003). *Sex differences in L2 vocabulary learning strategies*. *Applied Linguistics*, 13(1), 54–77. doi: 10.1111/1473-4192.00037.
- Chamot, A. U., & O'Malley, M. (1987). The cognitive academic language learning approach: A Bridge to the mainstream. *TESOL Quarterly*, 21(2), 227–250. doi: 10.2307/3586733.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6), 1–36. Retrieved from <http://www.jstatsoft.org/v61/i06/>
- Coady, J. (1997). L2 vocabulary acquisition: A synthesis of the research. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 273–290). Cambridge: Cambridge University Press.

- Cohen, A. D. (1998). *Strategies in using and learning a second language*. Essex: Addison Wesley.
- Cook, L. & Mayer, R. (1983). Reading strategies training for meaningful learning from prose. In M. Pressley & J. Levin (Eds.), *Cognitive Strategy Research: Educational Applications*, pp. 87–131. New York: Springer-Verlag.
- Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, 8, 103–128. doi: 10.1080/0958822940080202.
- Fillmore, L. W. (1983). The language learner as an individual: Implications of research on individual differences for the ESL teacher. *On TESOL*, 82, 157–173.
- Fu, I. (2005). *Students approaches to learning Chinese vocabulary*. Unpublished PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.
- Gu, Y. & Johnson, R. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643–679. doi: 10.1111/j.1467-1770.1996.tb01355.x.
- Gu, Y. (2003). Vocabulary learning in second language: Person, task, context and strategies. *Electronic Journal. TESL-EJ*, 7(2), 1–26.
- Hamzah, M. S. G., Kafipour, R., & Abdullah, S. K. (2009). Vocabulary learning strategies of Iranian undergraduate EFL students and its relation to their vocabulary size. *European Journal of Social Sciences*, 11(1), 39–50.
- Hindmarsh, R. (1980). *Cambridge English Lexicon*. Cambridge: Cambridge University Press.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689–696.
- Hulstijn, J. H. (2001). Intentional and incidental vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (ed.), *Cognition and second language instruction* (pp. 258–286). New York: Cambridge University Press.
- Kilgarriff, A. (2006). *BNC database and word frequency lists*. Retrieved from <http://www.kilgarriff.co.uk/bnc-readme.html#lemmatised>
- Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4(5), 900–913. doi: 10.4304/jltr.4.5.900-913.
- Kojic-Sabo, I., & Lightbown, P. (1999). Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal*, 83(2), 176–192. doi: 10.1111/0026-7902.00014.
- Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Kolb, D.A. (1985). *Learning style inventory*. Boston, MA: McBer.
- Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin*, 133, 464–481. doi: 10.1037/0033-2909.133.3.464.

- Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In M. McCarthy & N. Schmitt (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25(2), 21–33. doi: 10.1177/003368829402500202.
- Lessard-Clouston, M. (1996). ESL vocabulary learning in a TOEFL preparation class: A case study. *Canadian Modern Language Review*, 53, 97–119. doi: 10.3138/cmlr.53.1.97.
- Masrai, A. M. (2015). *Investigating and Explaining the Relationship between L1 Mental Lexicon Size and Organisation and L2 Vocabulary Development*. Unpublished PhD Thesis, Swansea University, Swansea.
- Masrai, A. M., & Milton, J. (2015). An investigation of the relationship between L1 lexical translation equivalence and L2 vocabulary acquisition. *International Journal of English Linguistics*, 5(2), 1–7. doi: 10.5539/ijel.v5n2p1.
- Masrai, A. M. (2016). The influence of morphological knowledge on lexical processing and acquisition: The case of Arab EFL learners. *Ampersand*, 3, 52–60.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- McCarthy, M. (2001). An interview with Professor Michael McCarthy. *Cambridge Connection*, 2(1). Hong Kong: China.
- Meara, P. & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (eds.) *Applied Linguistics in Society*. London: Centre for Information on Language Teaching and Research, vol. 24, no. 4, pp. 80–87.
- Meara, P. & Milton, J. (2003) *X_Lex*, The Swansea Levels Test. Newbury: Express.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Mondria, J. A. (2003). The effects of inferring, verifying and memorising on the retention of L2 word meanings. *Studies in Second Language Acquisition*, 25, 473–499.
- Nation, I. S. P. (1984). *Vocabulary lists: Words, affixes and stems*. English University of Wellington, New Zealand: English Language Institute.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle Publishers.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82.
- O'Malley, J. & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Ormrod, J. E. (2008). *Educational psychology: Developing learners* (6th ed.). Upper Saddle River, NJ: Pearson.

- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Englewood Cliffs, NJ: Newbury House.
- Oxford, R. L. (2003). *Language learning styles and strategies: An overview*. GALA. Retrieved from <http://web.ntpu.edu.tw/~language/workshop/read2.pdf>
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105–119. doi: 10.1111/j.1539-6053.2009.01038.x.
- Read, J. (1997). Vocabulary and testing. In N. Schmitt & M. McCarthy (eds.), *Vocabulary description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77–89. doi: 10.2307/3585941.
- Ruutemets, K. (2005). *Vocabulary learning strategies in studying English as a foreign language*. Master's thesis. Retrieved from <http://www.utlib.ee/ekollect/diss/mag/2005/b17557100/ruutemets.pdf>
- Safian, N. H., Malakar, S., & Kalajahi, S. A. R. (2014). Exploring vocabulary learning strategies used by UPM TESL undergraduates. *Advances in Language and Literary Studies*, 5(5), 1–4.
- Sanaoui, R. (1992). *Vocabulary learning and teaching in French as a second language classrooms*. Unpublished PhD Thesis, University of Toronto, Canada.
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt, and M. McCarthy, M. (eds.) *Vocabulary: Descriptive, Acquisition and Pedagogy*, pp. 199–227. Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Shangarfam, N., Ghorbani, N., Safarpour, E., & Maha, M. (2013). The comparative impact of lexical translation and lexical inferencing on EFL learners' vocabulary retention. *Management Science Letters*, 3(12), 2889–2908.
- Shen, Z. (2008). The roles of depth and breadth of vocabulary knowledge in EFL reading performance. *Asian Social Science*, 4, 135–137.
- Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental mental approaches to second language acquisition*, (pp. 81–94). Oxford: Pergamon.
- Sternberg, R. J., Grigorenko, E. L., & Zhang, L. (2008). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science*, 3, 486–506. doi: 10.1111/j.1745-6924.2008.00095.x.
- Stern, H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review*, 31(304), 318.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning*, 36, 139–152. doi: 10.1080/09571730802389975.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(04), 577–607.

- Stoffer, I. (1995). *University foreign language students' choice of vocabulary learning strategies as related to individual difference variables*. Unpublished Doctoral Dissertation, University of Alabama.
- Tanyer, S. & Ozturk, Y. (2014). Pre-service English Teachers' vocabulary learning strategy use and vocabulary size: A cross-sectional evaluation. *Journal of Language Teaching and Research*, 5(1), 37–45. doi: 10.4304/jltr.5.1.37-45.
- Wenden, A. L. (1982). Learner strategies. *TESOL Newsletter*, 19(5), 1–7.
- Wilkins, D. (1972). *Linguistics and language teaching*. London: Edward Arnold.

An Experimental Investigation of HAM, a Novel Mnemonic Technique for Learning L2 Homonyms and Homophones

Charles M. Mueller

Fuji Women's University, Sapporo, Japan

<https://doi.org/10.7820/vli.v07.1.mueller>

Abstract

Over the past 40 years, extensive research has examined the effectiveness of mnemonics for vocabulary learning. Much of this research has investigated the keyword method (Atkinson & Raugh, 1975), which involves linking an image related to a to-be-learned L2 word with an image related to a similar-sounding L1 word. Whereas most research has shown the keyword method to be effective (Webb & Nation, 2017) with impressive long-term retention rates (Beaton, Gruneberg, & Ellis, 1995), some have questioned its usefulness, particularly due to the quality of the resulting lexical representations and extended latencies associated with recall (Barcroft, Sommers, & Sunderman, 2011; Van Hell & Candia Mahn, 1997). Other drawbacks of the keyword technique are the equating of dissimilar L1 and L2 phonemes and the difficulty in creating associations for languages with markedly different phoneme inventories. The current study presents a novel approach called the Homonym/Homophone Association Method (HAM). It overcomes some of the drawbacks of the keyword method by associating meanings of L2 homonyms or homophones, one known by the learner and one unknown. Because the pronunciations of the L2 target words are identical (or nearly identical), learners only need to associate two distinct meanings. A quasi-experiment ($N = 71$) employing a within-subjects design compared the effectiveness of (1) HAM using researcher-generated associations and images, (2) HAM using self-generated associations, and (3) production practice that involved writing target words in sentences. Results on an unannounced posttest given 3 weeks after instruction suggest an advantage for HAM using researcher-generated associations.

Keywords: HAM; keyword method; homonyms; homophones; vocabulary learning; mnemonics; puns; humor

1. An Experimental Investigation of HAM, a Novel Mnemonic Technique for Learning L2 Homonyms and Homophones

In SLA research, deliberate vocabulary learning has not always received the attention it deserves (Boyd Zimmerman, 1997). This is unfortunate as a “well-chosen basic vocabulary” is essential to successfully complete everyday tasks in an L2 (second language) environment (De Groot & Van Hell, 2005, p. 9) and is highly predictive of L2 acquisition (Laufer, 1997).

At advanced levels of proficiency, vocabulary learning presents a significant challenge. For example, a learner of English may get by with knowledge of 4000 to 5000 word families, but knowledge of around 9000 word families is thought to be optimal for learners who need to independently read authentic English texts (Nation, 2015). To get a sense of the learning difficulty this “optimal” lexical knowledge poses, consider learning through extensive reading. It is estimated that learners must encounter a word around a dozen times to learn it through incidental exposure (Nation, 2014). Based on this, Nation (2014) predicts that a corpus of 2 427 807 tokens (roughly the number of words in 20 novels) would be needed for a learner to encounter the first 8000 word families of English an average of 12 times.

To assist learners with the formidable task of L2 lexical acquisition, incidental learning through massive amounts of input is ideally supplemented by deliberate learning (Webb & Nation, 2017). Language instructors are thus in need of pedagogical techniques to establish initial form-meaning links, especially as these relate to words in the fourth to ninth most frequent 1000 word families of English (i.e., “mid-frequency” words in Schmitt & Schmitt, 2014).

Over the last 40 years, researchers interested in investigating approaches to deliberate vocabulary learning have conducted extensive research on mnemonics. These memory techniques have been employed widely in education (for an overview and discussion, see Putnam, 2015). For example, Japanese school children, when learning a historical fact such as the date when Christopher Columbus arrived in the Americas, often associate the phrase *iyokuni moeru* (“[He] was burning with passion”) with Columbus since the first four syllables sound like the Japanese words for 1, 4, 9, and 2 (cp. Fontana, Scruggs, & Mastropieri, 2007).

In SLA, a large portion of mnemonics-related research has investigated the keyword method (Atkinson & Raugh, 1975; Pressley, Levin, & Delaney, 1982), which involves linking a target L2 word to a similar-sounding L1 word via an image incorporating both words. For example, an English L1 speaker who wants to memorize the Japanese word for cat (i.e., *neko*) might create an image of a man with a cat clinging to his neck. Research has generally shown the keyword method to be effective (Campos, Rodríguez-Pinal, & Pérez-Fabello, 2014) for L1 vocabulary learning, particularly for children with learning disabilities (Urberti, Scruggs, & Mastropieri, 2003). Applied to second language acquisition, this method has been shown to result in impressive long-term retention rates (Beaton et al., 1995), outperforming conventional deliberate vocabulary learning activities by around 20% (Webb & Nation, 2017, p. 115), although some studies (e.g., Hall, Wilson, & Patterson, 1981; Wei, 2015) have failed to show such advantages. Moreover, it has been found to be effective for learning a wide range of L2s, including English (Avila & Sadoski, 1996; Elhelou, 1994; Rodríguez & Sadowski, 2000), Galician (Campos et al., 2014), German (Desrochers, Wieland, & Coté, 1991; Fritz, Morris, Acton, Voelkel, & Etkind, 2007), Korean (Griffith, 1981), Russian (Atkinson & Raugh, 1975), Spanish (Gruneberg & Pascoe, 1996), and Tagalog (Wang, Thomas, & Ouellette, 1992). While most of the keyword studies have tested the method’s effectiveness for college-aged adults, it has also been found to be useful for early (Elhelou, 1994; Pressley, 1977) and older L2 learners (Gruneberg & Pascoe, 1996).

Even so, some research has questioned the suitability of the keyword method on various grounds. A key concern stems from findings that suggest that the quality of the resulting lexical representations is poor. A related concern is that the encoding of extra information as a cue (i.e., the imagery linking the two words) appears to result in extended latencies associated with recall (Barcroft et al., 2011; Van Hell & Candia Mahn, 1997). An additional drawback of the keyword technique is that it involves equating dissimilar L2 and L1 phonemes. A Japanese learner, for example, may choose to learn the English word *tar* by imagining a *taru* (the Japanese word for barrel) full of tar. Yet, this may inadvertently reinforce the Japanese learner's tendency to equate the English /r/ sound with the Japanese intervocalic [r] sound.

At a practical level, the keyword method is difficult to apply in the case of languages in which the L1 phoneme inventory and phonotactic constraints differ markedly from those of the L2 (cp. comments about difficulty of applying the keyword method with monosyllabic L1s in Webb & Nation, 2017, p. 116). For example, Japanese learners using the keyword method to learn English may find it difficult to learn English words with phonemes that are not present in Japanese. English learners may also find it difficult to apply the method to languages such as Japanese which, unlike English, contains many open vowels and unfamiliar phonemic contrasts such as vowel length or pitch accent.

The current study presents a novel mnemonic approach to vocabulary learning called the Homonym/Homophone Association Method (HAM). Languages typically contain words with distinct meanings that share the same written and/or spoken form. Words with unrelated meanings that share both spoken and written forms are commonly referred to as *homonyms*. An example would be *bill* (something you pay) and *bill* (a duck's beak). In many cases, homonyms actually have an etymological relationship that has become lost to language users (and is probably even less accessible for L2 learners). Since the current focus is on pedagogical approaches in L2 acquisition, words will be regarded as homonyms if the meaning relationship is not likely to be accessible to the L2 learner. In reality, the boundary separating homonyms (words with unrelated meanings) from polysemes (words with related meanings) is often fuzzy and scalar (for a practical measure of degrees of semantic relatedness, see Nagy & Anderson, 1984). Unlike homonyms, homophones (words like *hair* and *hare*) share the same sound but have different spellings. In some languages such as Japanese, the distinction between homonyms and homographs is also fuzzy. For example, many words are homonyms when written in *hiragana* (a phonetic script) but can be considered homographs when written in *kanji* (Sino-Japanese characters).

HAM is designed as an approach to L2 learning of homonyms and homophones, although it could, with some adaptations, perhaps be applied to homographs as well. While more limited in potential applicability, it overcomes some of the drawbacks of the keyword method by associating two L2 homonyms or homophones, one of which is known by the learner and one of which is unknown. Because the pronunciations of the L2 target words are identical, learners only need to associate two distinct meanings.

As an example of the technique, consider a learner of Japanese who wants to remember the low-frequency Japanese word "salmon roe" (*ikura*). The learner, even at an elementary stage of acquisition, is likely to know the homonym (*ikura*)

that means “how much”. Although the words are often distinct in Japanese writing, they are pronounced the same. Consequently, to learn the low-frequency meaning of *ikura*, the L2 learner could form a picture of someone in a market holding up a carton of salmon eggs while asking how much they cost (*Ikurawa ikura desuka?*)

Instructors creating pedagogical materials for instruction using HAM can readily find extensive materials from Internet sites (particularly, social networking sites) featuring humorous pictures. A common genre of humor on the Internet involves images in which a low-frequency member of a homonym or homophone pair appears within a context in which the high-frequency member would be expected. For example, in one such comic, an image shows an elderly man and woman with the woman saying, “You look quite distinguished with that grey hare on your head, Jim.” In the picture, in place of the man’s gray hair (expected based on context), there is a gray hare. This use of irony in humor can be explained in terms of Giora’s (1997) graded salience hypothesis, which maintains that more salient meanings of words (e.g., meanings that are more prototypical, familiar, frequent, conventional, and so on) will be retrieved from the lexicon more rapidly so that they temporally precede the processing of ironic interpretations. The hypothesis has received support in experiments that have compared reading times or response times of literal or ironic meaning (Giora & Fein, 1999; Giora, Fein, & Schwartz, 1998).

The idea of associating homonyms and homophones to aid memory is not new. It has been used in general education, for example, to recall definitions of key terms in history classes (Fontana et al., 2007). However, it has not received systematic treatment in SLA. Webb and Nation (2017), in their discussion of the keyword method, briefly allude to the possibility of associating the target word with L3 words (i.e., words from other foreign languages) or with other L2 words, yet they do not explicitly discuss homonyms or homographs.

In SLA, much of the research on the pedagogical potential of homonyms and homographs has focused on puns. For example, Lucas (2005) examined learner engagement with comic strips that featured puns. She found that the task focused learners’ attention on the phonological, morphological, and syntactic aspects of language and thus led to an increase in comprehension of the target meanings. Tocalli-Beller and Swain (2007), in a longer study, similarly demonstrated that pair-work in which L2 learners shared and explained riddles that involved homonyms led to sizable and durable gains in knowledge of the target lexical items. Bell (2012), in a qualitative study of spontaneous humorous interaction in an L2 classroom, concluded that humor facilitated the recall of language items, particularly in the case of word meanings. While these studies suggest that humor promotes the establishment of enduring form-meaning links, they did not explore whether the focus on homonyms (in the form of puns) promoted learning more effectively than alternative tasks.

In sum, previous research on homonyms and homographs has mostly focused on language play within classroom contexts, whereas research on mnemonics has focused on associations between L1 and L2 forms. The mnemonic potential of L2 homonyms and homographs has remained largely unexplored. This is unfortunate as the HAM technique would appear to have certain advantages. As with the keyword method, it should be particularly useful as a means

of forming initial form-meaning links, especially for words that are unlikely to be encountered frequently enough to be initially learned or retained. The use of a mnemonic should also help learners more readily recall a target word so that it can be used in production. By explicitly focusing on the phonological or written similarity of the to-be-learned word and a known word, HAM should also lead to more detailed encoding of the formal aspects of words. Finally, the method, by limiting its scope to equivalences between L2 word forms, completely avoids the potential problem of equating L2 and L1 phonemes (a possible objection to the keyword method). Finally, it adheres to recommendations that new words be learned through connections with pre-existing knowledge (Sökmen, 1997).

To determine whether HAM is an effective elaborative technique for teaching low-frequency vocabulary, the current study reports the results of an experiment comparing this approach with a more conventional vocabulary learning technique. The technique chosen for the comparison condition involved elaboration of an L2 item by using the target word in an English sentence. Boers (2015) points out that this word memorization technique has the advantage of stimulating learner engagement with both the form and meaning of the target word. Research has generally shown that writing a sentence with an L2 word is an effective means of learning the word as long as the word form is clearly linked to the associated meaning during production (cp. Barcroft, 2006). Coomber, Ramstad, and Sheets (1986), for example, found that college-aged learners of 10 artificial words who learned the words through sentence composition showed superior retention relative to participants who learned through rehearsal of definitions or exposure to examples (see also Zou, 2017).

In the current experiment, HAM instruction was further divided into two types to determine whether learners benefitted more from self-generated associations or instructor-generated associations. It is known that cognitive effort during encoding generally has a positive effect on recall (Tyler, Hertel, McCallum, & Ellis, 1979). While this would suggest that learners who generate their own keyword associations should show better retention, several studies that have examined the keyword method have failed to find significant differences between the two conditions (e.g., Colón & Rodriguez, 2004–2005; Hall, 1988; Shapiro & Waters, 2005). One possible explanation for the lack of a facilitative effect is that the benefits of effortful encoding may be offset by the time spent in generating (versus memorizing) the keyword mnemonic. The experiments in the current study explore whether this is also the case when the mnemonic technique targets L2 homonyms and homographs.

2. Method

A quasi-experiment employing a within-subjects design compared the effectiveness of (1) HAM using researcher-generated associations and images, (2) HAM using self-generated associations, and (3) production practice involving use of the target words in sentences.

Participants. The participants ($N = 71$) were first-year Japanese-L1 EFL learners from three intact classes at two universities in Japan. One class ($n = 16$) consisted of female students in a department focusing on linguistics and

literature at a private women's university. Those in the other two classes ($n = 27$, $n = 28$) were male and female students from a diverse range of majors who were first-year students at a large public university. Based on subjective impressions of participants' performance in class, they were, in terms of the Common European Framework of Reference (Council of Europe, 2001), mostly at Level B1 (threshold or intermediate), although some may have been at Level A2 (waystage or elementary).

Materials. The experiment targeted 30 words from various parts of speech. To avoid ceiling effects on the pretest, the senses of the targeted homonyms and homophones were low-frequency lexical items that the participants were unlikely to know.

The materials for the researcher-generated HAM (hereafter, "HAM Given") condition consisted of an 11-page hand-out. All instructions were provided in Japanese. The first page introduced the Homonym Association Method (HAM). The initial example showed a toad seated like a human being (i.e., with legs hanging down in front) on a bench at a bus stop with the caption: "Just waiting for the bus cause my car got toad". The text explained that the humor of the text involved the replacement of *towed* (the word that would naturally come to mind in this situation) with the homophonous word *toad*, and the ludicrous notion that the toad would drive or wait for a bus. The following 10 pages had similar humorous pictures associating two L2 words. In the materials for this condition and the other two conditions (discussed below), four of the ten target words involved homonyms, three involved homophones, and three involved "near" homonym/homophone associations in which both the spelling and spoken form were slightly different (e.g., *dessert* and *deserted*). Below the picture on each page, the associated L2 words were given with their Japanese translations. Below these words, there were seven lines. Participants were asked to explain in Japanese why the picture was humorous, modeling their responses on the explanation on the first page of the hand-out (the page with the toad at the bus stop).

The materials for the HAM Self-Generated condition introduced the method in Japanese using the example discussed in the introduction (i.e., the association of the two meanings of *ikura* using an image of a shopper asking how much salmon roe costs at a fish market). In addition to this example, targeting the association of homonyms, an example was provided of a homophone association in which the words *die* and *dye* were linked by means of a humorous picture showing an Easter egg dipped in dye saying, "I don't want to die in this dye!" It was pointed out that HAM could also be used when there was a slight mismatch between the two words being associated (as is the case, for example, with *dessert* and *deserted*). The following pages of the hand-out showed two to-be-associated words with their Japanese translations. Below the two words were blank lines. Participants were asked to think of an image associating the two words and then write a sentence linking the two words, modeling their responses on the examples provided on the first page (i.e., the *ikura* example and the *die/dye* example).

In the Sentence Composition condition, participants were given an English word with a Japanese translation. Below the words, there were blank lines where participants were to write a sentence using the targeted sense of the English word.

To ensure that the relative difficulty of the words would not bias the results, the three sets of 10 words were counterbalanced so that each set appeared in a different condition for each class. The pretest and posttest measures were identical and consisted of a translation task in which the L1 word had to be translated into the L2 target. The test consisted of 30 items ($\alpha = .77$). To prevent participants from translating the L1 word with an untargeted L2 word, the number of letters of the target word (shown with bolded underlines) and some of the letters were provided.

Procedure. The experiment was conducted in participants' regular class during regular class times with their regular instructor (the author). Participants received the pretest forms and were given 20 minutes to take the test. After completing the test, they learned the 30 target words, 10 words at a time, as they completed learning tasks in each of the three conditions. To ensure that sequencing of the conditions did not influence results, the order was counterbalanced so that the three classes did the three conditions in different sequences. Thus, the (A) HAM Given condition, (B) the HAM Self-Generated condition, and (C) the Sentence Composition condition were performed in differing sequences (i.e., ABC, BCA, and CAB) by each class.

Participants received hand-outs for one of the experimental conditions, and the instructor briefly went over the directions with the class. The participants then had 20 minutes to complete the task for that condition. The same procedure was followed for the other two conditions. Three weeks after the pretest and intervention, participants were given an unannounced posttest during which they were given 20 minutes to answer the questions. It should be noted that there was no immediate posttest and instead, the "posttest" was given later, at the point when a delayed posttest is typically given in SLA experiments. This format was adopted to avoid the confounding practice effects that can occur when an immediate posttest precedes a delayed posttest (for another methodological approach to this issue, see Avila & Sadoski, 1996; Wang et al., 1992).

Scoring. To ensure that the test measure captured partial learning, participants were given credit for the percentage of each answer that was correct. For example, the test item *hoarse* appeared on the test with five blanks (the "s" was already provided). Since there were five blank spaces for that item, a correct letter occurring at each correct position was counted as worth 20% of the total points for that item.

3. Results

The participants appeared to have ample time to complete the pretest and posttest, thus time pressure was unlikely to have affected the results. The pretest and posttest scores for the three conditions are shown in Table 1. For all tests, the maximum possible score was 30 points.

As can be seen from the pretest results, the participants ($N = 71$) knew only around 10% of the target words prior to the intervention. On the posttest, the scores for the HAM Self-Generated and Sentence Composition conditions were virtually identical at just over 20%, whereas the HAM Given condition led to scores just

Table 1. Pretest and Posttest Scores for the Three Within-Subjects Conditions

HAM given		HAM self-generated		Sentence composition	
Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Range	Range	Range	Range	Range	Range
3.35 (2.6)	8.89 (6.0)	2.44 (2.3)	6.48 (4.2)	3.13 (2.9)	6.78 (4.3)
0.0–9.7	0.0–22.8	0.0–7.7	0.0–21.2	0.0–12.0	0.0–18.0

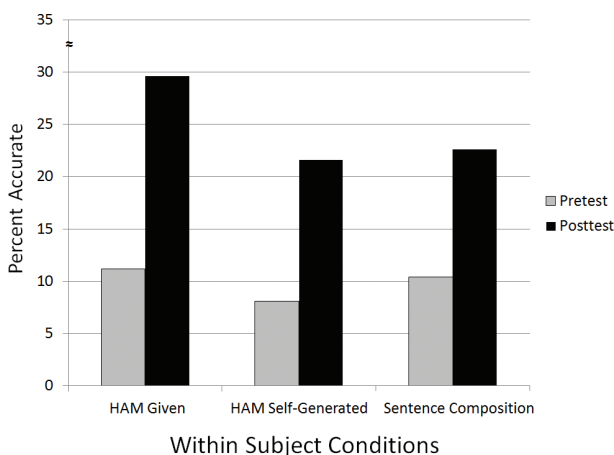


Figure 1. Percentage of Items Answered Correctly on the Pretest and Posttest for each of the Three Within-Subjects Conditions.

under 30%. The higher SDs on the HAM Given posttest suggests greater variability in participants' ability to take advantage of this type of instruction. Figure 1 shows the same within-subject scores in terms of percentage accurate.

To determine whether the differences between scores in the three conditions were statistically significant, an ANCOVA was conducted with Type of Instruction as the independent variable with three levels (HAM Given, Ham Self-Generated, and Sentence Composition), with posttest scores as the dependent variable, and with pretest scores as the covariate. An alpha of .05 was used for all tests. The results of the ANCOVA indicated that there were significant differences among the three adjusted means, $F(2,209) = 4.00$, $p = .020$, $\eta_p^2 = .037$. Post hoc tests were conducted to determine which conditions differed. The confidence intervals were all adjusted for multiple comparisons using a Sidak correction. Scores on target words learned in the HAM Given condition were significantly higher than those for words learned in the Sentence Composition condition ($p = .025$, 95% CI [0.6, 12.1]), but the difference between the HAM Given and HAM Self-Generated scores fell short of statistical significance ($p = .090$, 95% CI [-0.6, 11.0]). The negligible difference between the HAM Self-Generated and Sentence Composition scores was not significant ($p = .956$).

Partial scoring had only a minor effect on the results. Most responses were 100% accurate or were completely inaccurate, in the latter case, usually due to

participants' failure to provide a response. Fewer than 7% of participants' responses involved partial scores greater than zero and less than 75%. (It should be noted that partial scores over 75% usually reflected minor spelling errors).

4. Discussion

The current experiment sought to determine whether a mnemonic vocabulary learning technique using either researcher-generated or self-generated images to associate the different senses of homonyms and homophones (i.e., HAM) would be more effective than Sentence Composition, a conventional vocabulary technique used to promote elaborative encoding of form-meaning associations. The results suggest that HAM was more effective than Sentence Composition when the associations and images were provided to participants. When participants had to generate their own images and associations, HAM and Sentence Composition appeared to be equally effective.

In the current study, the locus of the effect is not entirely clear. A straightforward explanation would be that HAM, like the keyword method, is effective due to the facilitative effects of imagery on lexical acquisition (Campos, Amor, & González, 2004; Farley, Ramonda, & Liu, 2012; Paivio, 1986). In the current experiment, participants achieved more learning in the condition in which they were provided with images. It could be that these images were a crucial factor promoting greater learning in the HAM Given condition (cp. Thomas & Wang, 1996).

Another possibility is that the associations learned in the HAM Given condition were better retained due to humor. While the effects of humor on learning have been investigated within the general field of education (e.g., Martin, Preiss, Gayle, & Mike, 2006), they have only begun to be explored in the area of vocabulary acquisition (for an overview, see Bell, 2017). Bell (2012) found superior retention for lexical items when the incidental focus on the word involved humor. Along similar lines, some research on the keyword method (e.g., Campos, Amor, & González, 2002) has found that bizarre images facilitate learning more than conventional images. Future research on HAM and related mnemonic techniques may therefore gain greater insights if the presence of humor is treated as an independent variable manipulated within the research design.

In terms of practical implications, the results of the current study are quite positive, suggesting that HAM provides a highly effective alternative to conventional vocabulary instruction techniques. It should be noted that pedagogical applications of the technique will require some adjustments. In the current experiment, the key criterion for selecting target words was that they would be unknown to participants prior to the experiment; hence, most of these words would not be useful to typical students learning English. When applied to actual pedagogical situations, the selection would need to be motivated by students' needs. HAM is likely to be particularly effective when students have good knowledge of one sense of a homonym and can benefit from learning a different sense of the same homonym, or when one member of a homophone pair (e.g., *hair*) is known and the other (e.g., *hare*) unknown. While vocabulary acquisition in the current experiment was fairly modest (an increase from just over 11% to nearly 30% for the HAM Given condition), it should be kept in mind that long-term retention of vocabulary

from a single brief exposure is rare. Typically, deliberate learning of words requires around seven repetitions to result in enduring form-meaning associations (Webb & Nation, 2017). Viewed in this light, the durable learning for words studied in the HAM Given condition in the current experiment was quite remarkable. These retention rates could probably be even further enhanced by combining HAM with some additional techniques such as flashcard review of target words using spaced repetition (Griffin & Harley, 1996; Roediger & Karpicke, 2010).

It must be acknowledged that the homonym association method, being applicable primarily to homonyms and homophones (and to a lesser extent, to near-homonyms, near-homophones, and homographs) is of much more limited scope than the keyword method, which can be used on a large range of words. That said, homonyms and homophones are quite frequent in many languages. According to Parent (2012), 3.3% of English words (75 words) on the roughly 2000-word General Service List (West, 1953) are homonyms, whereas 6.4% of the words (147 words) are homophones. Parent furthermore estimates that around 10% of the 2284 most frequent words in English are homonyms, homophones, or homographs. Wang Ming-tzu and Nation (2004) likewise claim that around 10% of the Academic Word List (Coxhead, 2000) consists of homonyms and homographs.

Future research will need to verify the effects for HAM observed in the current study. Although time on task was controlled, it is not clear whether HAM would be superior to other alternatives such as a word association condition (e.g., rote memorization of L2–L1 pairs), especially if the word pairs were presented multiple times at a fast rate (cp. Hall, Owens, & Wilson, 1987). Although multiple exposures to words within a short session lead to diminishing returns, the gains are usually superior to single exposures (Nakata, 2017). It could be that a paired association learning condition involving multiple retrievals would be more effective than HAM if time on task were controlled or if learning were operationalized in terms of items acquired per fixed time interval (e.g., words successful learned per minute of study). It could also be that the HAM would be even more effective if combined with other conventional techniques. Future research should therefore examine HAM in comparison with (and in combination with) other deliberate learning approaches.

The current study found that self-generated mnemonics were markedly inferior to mnemonics given to the participants. Some general educational research suggests that mnemonics are more effective when they are generated by learners if the learners are adequately trained in the method and have sufficient time during encoding (Putnam, 2015). The current study strictly controlled time on task, so it is possible that participants were unable to generate effective mnemonic associations within the allotted time. The current results are actually in line with previous SLA research on the keyword method, which has found that conditions in which learners generate their own associations lead to similar or less learning than conditions in which the associations are provided (Campos et al., 2002).

Future research should also confirm whether HAM results in representations that are qualitatively inferior (cp. Barcroft et al., 2011) to those developed using other methods. The issue of lexical representation is complex, as there are

currently a number of competing models (Heredia & Brown, 2013). Some models assume that a bilingual's languages are stored separately but are linked to a common conceptual system. Based on this assumption, the Revised Hierarchical Model (Kroll & Stewart, 1994) further claims that lexical links from L2 to L1 words are established prior to the formation of strong semantic links between L2 words and their corresponding concepts. L1 words, on the contrary, have been shown to be strongly linked to their corresponding concepts. This asymmetry, typical of less proficient learners, is said to be especially strong in the case of production (Kroll, Van Hell, Tokowicz, & Green, 2010). L2 representations could also be examined based on distributional models (e.g., Van Hell & de Groot, 1998).

While a detailed discussion of models of bilingual lexical representation is beyond the scope of the current paper, one fruitful avenue of future research would be to use theory-based psycholinguistic behavioral measures to explore the type of representations created through HAM- versus keyword-based instruction. Ideally, such research would take into account learning over long time intervals. Like the keyword method, HAM would presumably lead to representations that are initially mediated by extraneous information (i.e., the image associating the two senses of the homonym or homograph), and this mediation would be expected to slow recall. In the long term, however, HAM and the conventional keyword method may lead to more enduring initial form-meaning links that allow for the development of unmediated form-meaning links as the target word is subsequently encountered incidentally in input and the form-meaning link is processed.

Finally, HAM should, in future research, be compared directly with the conventional keyword method. The keyword method has been described by some researchers as being of "low utility" in practice due to drawbacks related to acceleration of forgetting and difficulty of application (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013, p. 24). It would be interesting to know if HAM provides some advantages in this regard.

References

- Atkinson, R.C., & Raugh, M.R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 1(2), 126–133. <https://doi.org/10.1037/0278-7393.1.2.126>
- Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, 46(3), 379–395. <https://doi.org/10.1111/j.1467-1770.1996.tb01241.x>
- Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research*, 22(4), 487–497. <https://doi.org/10.1191/0267658306sr276oa>
- Barcroft, J., Sommers, M.S., & Sunderman, G. (2011). Some costs of fooling Mother Nature: A priming study on the Keyword Method and the quality of developing L2 lexical representations. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research*:

- Insights from psycholinguistics* (pp. 49–72). Amsterdam, The Netherlands: John Benjamins.
- Beaton, A., Gruneberg, M.M., & Ellis, N.C. (1995). Retention of foreign vocabulary learned using the keyword method: A ten-year follow-up. *Second Language Research*, 11(2), 112–120. <https://doi.org/10.1177/026765839501100203>
- Bell, N.D. (2012). Comparing playful and nonplayful incidental attention to form. *Language Learning*, 62(1), 236–265. <https://doi.org/10.1111/j.1467-9922.2011.00630.x>
- Bell, N.D. (2017). Humor and second language development. In S. Attardo (Ed.), *The Routledge handbook of language and humor* (pp. 444–455). New York: Routledge.
- Boers, F. (2015). Words in second language learning and teaching. In J.R. Taylor (Ed.), *The Oxford handbook of the word* (pp. 582–596). Oxford, UK: Oxford University.
- Boyd Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 5–19). Cambridge, UK: Cambridge University Press.
- Campos, A., Amor, Á., & González, M.Á. (2002). Presentation of keywords by means of interactive drawings. *The Spanish Journal of Psychology*, 5(2), 102–109. <https://doi.org/10.1027/1618-3169.51.2.125>
- Campos, A., Amor, Á., & González, M.Á. (2004). The importance of the keyword-generation method in keyword mnemonics. *Experimental Psychology*, 51(2), 125–131. <https://doi.org/10.1027/1618-3169.51.2.125>
- Campos, A., Rodríguez-Pinal, M.D., & Pérez-Fabello, M.J. (2014). Receptive and productive recall with the keyword mnemonics in bilingual students. *Current Psychology*, 33(1), 64–72. <https://doi.org/10.1007/s12144-013-9197-y>
- Colón, M., & Rodríguez, M. (2004–2005). The effects of self-generated versus supplied interactive images on the long-term retention of EFL vocabulary by college students. *Lenguas Modernas*, 30, 191–200.
- Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge, UK: Cambridge University.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>.
- De Groot, A.M.B., & Van Hell, J.G. (2005). The learning of foreign language vocabulary. In J.F. Kroll & A.M.B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9–29). New York: Oxford University.
- Desrochers, A., Wieland, L.D., & Coté, M. (1991). Instructional effects in the use of the mnemonic keyword method for learning German nouns and their grammatical gender. *Applied Cognitive Psychology*, 5(1), 19–36. <https://doi.org/10.1002/acp.2350050103>.
- Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques: Promising

- directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Elhelou, M.-W.A. (1994). Arab children's use of the keyword method to learn English vocabulary words. *Educational Research*, 36(3), 295–302. <https://doi.org/10.1080/0013188940360308>
- Farley, A.P., Ramonda, K., & Liu, X. (2012). The concreteness effect and the bilingual lexicon: The impact of visual stimuli attachment on meaning recall of abstract L2 words. *Language Teaching Research*, 16(4), 449–466. <https://doi.org/10.1177/1362168812436910>
- Fontana, J.L., Scruggs, T.E., & Mastropieri, M.A. (2007). Mnemonic strategy instruction in inclusive secondary social studies classes. *Remedial and Special Education*, 28(6), 345–355. <https://doi.org/10.1177/07419325070280060401>
- Fritz, C.O., Morris, P.E., Acton, M., Voelkel, A.R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, 21(4), 499–526. <https://doi.org/10.1002/acp.1287>
- Giora, R. (1997). Understanding figurative and literal language. The graded salience hypothesis. *Cognitive Linguistics*, 7(1), 183–206. <https://doi.org/10.1515/cogl.1997.8.3.183>
- Giora, R., & Fein, O. (1999). Irony comprehension: The graded salience hypothesis. *Humor*, 12(4), 425–436. <https://doi.org/10.1515/humr.1999.12.4.425>
- Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Graded salience and indirect negation. *Metaphor and Symbol*, 13(2), 83–101. https://doi.org/10.1207/s15327868ms1302_1
- Griffith, D. (1981). An evaluation of the key-word technique for the acquisition of Korean vocabulary by military personnel. *Bulletin of the Psychonomic Society*, 17(1), 12–14. <https://doi.org/10.3758/BF03333652>
- Griffin, G., & Harley, T.A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17(4), 443–460. <https://doi.org/10.1017/S0142716400008195>
- Gruneberg, M.M., & Pascoe, K. (1996). The effectiveness of the keyword method for receptive and productive foreign vocabulary learning in the elderly. *Contemporary Educational Psychology*, 21(1), 102–109. <https://doi.org/10.1006/ceps.1996.0009>
- Hall, J.W. (1988). On the utility of the keyword mnemonic for vocabulary learning. *Journal of Educational Psychology*, 80(4), 554–562. <https://doi.org/10.1037/0022-0663.80.4.554>
- Hall, J.W., Owens, W.L., & Wilson, K.P. (1987). Presentation rates and keywords in vocabulary learning. *Bulletin of the Psychonomic Society*, 25(3), 179–181. <https://doi.org/10.3758/BF03330321>
- Hall, J.W., Wilson, K.P., & Patterson, R.J. (1981). Mnemotechnics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology*, 73(3), 345–357. <https://doi.org/10.1037/0022-0663.73.3.345>

- Heredia, R.R., & Brown, J.M. (2013). Bilingual memory. In T.K. Bhatia & W.C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 269–291). Malden, MA: Blackwell.
- Kroll, J.F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. <https://doi.org/10.1006/jmla.1994.1008>
- Kroll, J.F., Van Hell, J.G., Tokowicz, N., & Green, D.W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–381.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20–34). Cambridge: Cambridge University Press.
- Lucas, T. (2005). Language awareness and comprehension through puns among ESL learners. *Language Awareness*, *14*(4), 221–238. <https://doi.org/10.1080/09658410508668838>
- Martin, D.M., Preiss, R.W., Gayle, B.M., & Mike, A. (2006). A meta-analytic assessment of the effect of humorous lectures on learning. In B.M. Gayle, R.W. Preiss, N. Burrell, & M. Allen (Eds.), *Classroom communication and instruction processes* (pp. 295–313). Mahwah, NJ: Erlbaum.
- Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*(3), 304–330. <https://doi.org/10.2307/747823>
- Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*, *39*(4), 653–679. <https://doi.org/10.1017/S0272263116000280>
- Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16.
- Nation, P. (2015). Which words do you need? In J.R. Taylor (Ed.), *The Oxford handbook of the word* (pp. 568–581). Oxford, UK: Oxford University.
- Paivio, A. (1986). *Mental representation: A dual coding approach*. New York: Oxford University.
- Parent, K. (2012). The most frequent English homonyms. *RELC Journal*, *43*(1), 69–81. <https://doi.org/10.1177/0033688212439356>
- Pressley, M. (1977). Children's use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology*, *69*(5), 465–472.
- Pressley, M., Levin, J.R., & Delaney, H.D. (1982). The mnemonic keyword method. *Review of Educational Research*, *52*(1), 61–91. <https://doi.org/10.3102/00346543052001061>
- Putnam, A.L. (2015). Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, *1*(2), 130–139. <https://doi.org/10.1037/tps0000023>

- Rodriguez, M., & Sadowski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, 50(2), 385–412. <https://doi.org/10.1111/0023-8333.00121>
- Roediger, H.L., III, & Karpicke, J.D. (2010). Intricacies of spaced retrieval: A resolution. In A.S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 23–47). New York: Psychology Press.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Shapiro, A.M., & Waters, D.L. (2005). An investigation of the cognitive processes underlying the keyword method of vocabulary learning. *Language Teaching Research*, 9(2), 129–146. <https://doi.org/10.1191/1362168805lr151oa>
- Sökmen, A.J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237–257). Cambridge, UK: Cambridge University.
- Thomas, M.H., & Wang, A.Y. (1996). Learning by the keyword mnemonic: Looking for long-term benefits. *Journal of Experimental Psychology: Applied*, 2(4), 330–342. <https://doi.org/10.1037/1076-898X.2.4.330>
- Tocalli-Beller, A., & Swain, M. (2007). Riddles and puns in the ESL classroom: Adults talk to learn. In A. Mackey (Ed.), *Conversational interaction and second language acquisition: Empirical studies* (pp. 143–167). Oxford, UK: Oxford University.
- Tyler, S.W., Hertel, P.T., McCallum, M.C., & Ellis, H.C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 607–617. <https://doi.org/10.1037/0278-7393.5.6.607>
- Urberti, H.Z., Scruggs, T.E., & Mastropieri, M.A. (2003). Keywords make the difference!: Mnemonic instruction in inclusive classrooms. *TEACHING Exceptional Children*, 35(3), 56–61. <https://doi.org/10.1177/004005990303500308>
- Van Hell, J.G., & Candia Mahn, A.C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, 47(3), 507–546. <https://doi.org/10.1111/0023-8333.00018>
- Van Hell, J.G., & de Groot, A.M.B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193–211. <https://doi.org/10.1017/S1366728998000352>
- Wang, A.Y., Thomas, M.H., & Ouellette, J.A. (1992). The keyword mnemonic and retention of second language vocabulary. *Journal of Educational Psychology*, 84(4), 520–528. <https://doi.org/10.1037/0022-0663.84.4.520>
- Wang Ming-tzu, K., & Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics*, 25(3), 291–314. <https://doi.org/10.1093/applin/25.3.291>

- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University.
- Wei, Z. (2015). Does teaching mnemonics for vocabulary learning make a difference? Putting the keyword method and the word part technique to the test. *Language Teaching Research*, 19(1), 43–68. <https://doi.org/10.1177/1362168814541734>
- West, M. (1953). *A general service list of English words*. London, UK: Longman, Green and Co.
- Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, 21(1), 54–75. <https://doi.org/10.1177/1362168816652418>

Examining the Word Family through Word Lists

Dale Brown

Institute of Liberal Arts and Science, Kanazawa University, Kanazawa, Japan

doi: <https://doi.org/10.7820/vli.v07.1.brown>

Abstract

The choice of lexical unit has important consequences for L2 vocabulary research, testing and instruction. In recent years, the most widely used lexical unit has been the word family. This study examines the characteristics of word lists based on the word family and explores the levels of text coverage such lists may provide should the assumption that learners can deal with word families be incorrect. This is pursued through the detailed examination of a set of word-family-based word lists. The study finds that such word lists pose a number of challenges, including the number of word forms with multiple affixes, the number of word forms with more challenging affixes, and the number of word families in which the base word is not the most frequently occurring member. Moreover, the first thousand word families in particular are shown to be challenging. The study then demonstrates that if learners are unable to deal with the complexity of word families, even to a relatively small degree, word-family-based lists may provide far lower text coverage levels than may be assumed. It concludes that in work on second language vocabulary, careful consideration is needed of the appropriacy of the word family as the lexical unit and highlights the range of work based on the word family that may need reevaluating.

Keywords: lexical unit, word families, word lists, text coverage

1 Introduction

The choice of lexical unit has important consequences for work on second language vocabulary (Gardner, 2007), with implications for research, language testing, curriculum design, and teaching. In recent years, the word family has been the most widely used lexical unit. The word family, described by Bauer and Nation (1993), was intended as a flexible concept, with seven levels (see Table 1).

Despite the intention of flexibility, in practice Level 6 word families (i.e., including word forms featuring any of the Levels 1–6 affixes; referred to hereafter as WF6) have been used most often, and indeed, reference to “word families” in the literature largely refers to this specific level. This is primarily due to the free availability of word lists based on WF6 developed by Nation, but is also a result of the extensive work of Nation and colleagues featuring these word lists (e.g., Beglar, 2010; Laufer & Ravenhorst-Kalovski, 2010; Webb & Macalister, 2013), and the freely available Range (Nation & Heatley, 2002) software and its online equivalent at LexTutor (Cobb, no date) which make use of these lists.

Table 1. Summary of Bauer and Nation's (1993) Word Families Scheme

Level	Description	Number of affixes	Examples of affixes	Examples of forms
1	Each form is a different word.			
2	Inflectional suffixes.	8	<i>-ed, -s</i>	<i>heat-ed</i> <i>waste-s</i>
3	The most frequent and regular derivational affixes.	10	<i>-able, -er, non-</i> (each with restricted uses)	<i>heat-er</i>
4	Frequent, orthographically regular affixes.	11	<i>-al, -ful, in-</i> (each with restricted uses)	<i>waste-ful</i>
5	Regular but infrequent affixes.	50	<i>-age, -ally, ante-</i>	<i>wast-age</i>
6	Frequent but irregular affixes.	12	<i>-ee, -ic, pre-</i>	<i>pre-heat</i>
7	Classical roots and affixes.			

Alternatives to the word family include word types, lemmas, and flemmas. A word type is an individual word form. Thus, *act* and *acted* are different word types. A lemma is a base word and its inflections (i.e., paradigmatically related forms of the same word class). Thus, *act*^{verb} and *acted*^{verb} are part of a single lemma, while *act*^{noun} belongs to a different lemma. A flemma (Pinchbeck, 2014, March) is a base word and inflected forms regardless of word class. Thus, *act*^{verb}, *act*^{noun}, and *acted*^{verb} are part of a single flemma, while *actor* belongs to a different flemma.

Nation (2006b; 2015) argues that different lexical units are appropriate for different purposes, but suggests word families are a good choice when considering receptive uses of language, except with learners who are very beginners. Nation cites two justifications: first, there is evidence that word families are psychologically real (Bertram, Laine & Virkalla, 2000; Nagy, Anderson, Schommer, Scott & Stallman, 1989), and second, once learners have some familiarity with a word family, they are able to deal with its various members with little difficulty when encountered in context.

There has, however, been some questioning of the word family. It has been pointed out that the research cited by Nation in justifying the word family was with L1 participants (McLean, 2017), while the compilers of two recently developed word lists (Brezina & Gablasova, 2015; Gardner & Davies, 2014) expressed concerns about the semantic distance that can exist between individual forms in a word family and about whether learners have the morphological skills necessary to deal with derivational word relationships, leading both to choose lemmas as their lexical unit.

More pointedly, several studies have shown that L2 learners do not necessarily find dealing with the types of word forms that word families contain a simple task. Mochizuki and Aizawa (2000) presented Japanese L2 learners with pseudowords featuring affixes in Levels 3–6 of Bauer and Nation's scheme and, in a multiple-choice format, asked learners to choose the meaning of the affix in the case of prefixes and the part of speech of the word in the case of suffixes. Knowledge of affixes correlated with vocabulary size, but remained partial even among learners with a vocabulary size estimated at over 5000 words. Furthermore, knowledge

of particular affixes did not correspond with their position in Bauer and Nation's scheme (e.g., the best known prefix in the study was *re-*, yet this appears at Level 6 in the scheme). In a replication of Mochizuki and Aizawa's study with upper-intermediate Serbian learners, Danilović, Savić, and Dimitrijević (2013) likewise found partial knowledge of English affixes and an apparent order of acquisition at odds with Bauer and Nation's scheme. Ward and Chuenjundaeng (2009) had Thai learners of English give translations of base forms and of derived forms and found that in most cases where a base form was translated successfully, the derived form was not, and vice versa. They concluded that Thai learners do not in general make use of English word-building devices. Brown (2013) asked Japanese L2 learners to mark unknown words while reading and then investigated the characteristics of the marked words. This revealed that, in high-frequency word families, inflectional and derivational forms were relatively more likely to be marked than base words, suggesting such forms pose some additional difficulty. Reynolds (2015) conducted a study of incidental vocabulary learning with advanced learners in Taiwan and found better acquisition for words that occurred in the text in a single invariant form as compared with words whose occurrences displayed inflectional or derivational variation. Finally, McLean (2017), in a study with Japanese learners at various proficiency levels, tested the receptive understanding of a number of highly frequent words and multiple members of their word families. There was good comprehension of inflectional forms, but limited comprehension of derivational forms even among learners of advanced proficiency. There is, then, evidence, in the case of both learners with L1s that make use of derivation and those that do not, that learners up to advanced levels of proficiency find the degree of knowledge of derivations and word-building processes necessary to deal with the word family a considerable challenge.

The adoption of the word family as the lexical unit is not, however, without attractions: it enables higher levels of text coverage to be achieved with a smaller word list. Text coverage matters because it affects comprehension (Hu & Nation, 2000; Schmitt, Jiang & Grabe, 2011). One aim in developing word lists is therefore a desire to discover how higher coverage levels can be achieved most efficiently, that is, to identify what learners should learn in order to reach greater coverage levels as quickly and easily as possible. With an expanded concept of word, such as WF6, greater coverage can be achieved, and so word-family-based lists can make the vocabulary learning challenge appear relatively achievable and manageable.

The crucial issue is, however, whether learners can indeed deal with the various members of a word family when encountering them in context. If they cannot, the coverage levels that it is claimed can be reached are illusory. What is more, the assumption need only be somewhat incorrect for problems to arise. That is, learners who can deal with much of the challenge posed by word families, but not all, may nonetheless face problems. If, for example, learners can deal with 90% of what lies behind a word-family-based word list (i.e., learners have 90% of the knowledge needed to cope with word families), the true coverage level provided by that list for a given text may not be, say, 98%, but rather 88% (98% × 90%). Such a drop in coverage may seem insignificant, but small differences in coverage can substantially affect comprehension (Hu & Nation, 2000; Schmitt, Jiang & Grabe, 2011). This is easier to recognize by switching the perspective and

considering that with 98% coverage 2% of words are unknown, while with 88% coverage 12% are unknown, six times more.

What is absent from the debate about the word family thus far is a thorough interrogation of word-family-based word lists that would plainly reveal their characteristics. The affixes permitted are described by Bauer and Nation (1993), and actual lists themselves can be freely downloaded and examined (see, e.g., <http://www.victoria.ac.nz/lals/about/staff/paul-nation>). Yet, it is not easy to gain an understanding of the nature of the lists: they are simply too large, containing thousands of forms, for any casual perusal to be informative. It has been pointed out, for example, that the base word of a family can be less frequent than other members of the family (Coniam, 1999). However, it is not clear if this is an isolated instance or a more generalized problem. A more systematic examination of word families would provide teachers and researchers with a better sense of the challenge that learners face in dealing with word families and of the consequences should the assumption that learners can deal with this challenge be mistaken.

Thus, this study does not replicate those cited above and explore how learners deal with word families. Instead, the aim is to conduct a thorough examination of a set of word-family-based word lists in order to provide a detailed characterization of such lists. Specifically, two questions are asked:

1. To what extent do WF6-based word lists have characteristics that may be challenging for learners?
2. If the assumption that learners can deal with WF6 is incorrect, what levels of text coverage might such lists provide?

The first question involves investigating the size of the word families in the lists and their complexity; the second question means looking at the frequency of individual forms within the word families that include affixes at different levels in the word families scheme.

2 Method

This study analyzes the higher frequency portion of Nation's (2006a) British National Corpus-based word lists. These lists consist of fourteen bands, each of 1000 word families. This study examines the first five bands, so as to concentrate on the bands that contain the vast majority of words in any text and which are the focus of vocabulary learning for the majority of L2 learners (Webb & Sasao, 2013).

Nation's BNC-based lists contain WF6 word families as established by Bauer and Nation. However, the lists also include some forms featuring affixes outside of the scheme (e.g., several forms including the affix *dis-* as in *dislike*; see the *Other affixes observed* column in the additional material online), along with irregular verb and noun forms (e.g., *became* within the BECOME family), abbreviated forms of base words (e.g., *ad* and *advert* within the ADVERTISE family), alternative spellings of base words (e.g., *center* within the CENTRE family), and compound forms (e.g., *backbone* within the BACK family).

The BNC-based lists were selected for study since they are based on a single source which can be easily accessed. This means that the frequency of the forms

in the lists can be checked in the very source of the lists itself. The lists were downloaded from Nation's website (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>), and all searches of the BNC were conducted via the BYU-BNC (Davies, 2004) website (<http://corpus.byu.edu/bnc>). For the majority of the analyses, all 1000 word families in each of the five bands were examined. However, for the corpus-based investigation of the lists, a systematic random sample of 100 word families was taken from each band.

Frequency information was collected for each of the 2396 word forms in the five samples of 100 word families. In each case, the search was for the word form itself (i.e., no part of speech was specified).

In order to address research question 1, on the characteristics of WF6-based word lists, the analysis looks at:

- the number of word forms in the families across the five bands
- the number of word forms in each band that contain different numbers of affixes (i.e., the number of forms including a single affix, the number of forms including two affixes, and so on)
- the number of word forms in each band that include affixes at the various levels of Bauer and Nation's scheme (i.e., the number of forms including Level 2 affixes only, the number of forms including affixes through to Level 3, and so on)
- the number of word families for which the base word is not the most frequent member of the family.

The analysis then addresses research question 2, on text coverage levels. This was a two-step process:

1. Based on the frequency in the BNC of the individual forms that comprise word families, a calculation was made of the mean proportion of a word family's total occurrences that is provided by forms at different levels of the scheme (i.e., the proportion of a word family's total occurrences accounted for by forms including Level 2 affixes only, by forms through to Level 3, and so on).
2. The above proportions were then used to make estimates of the varying degrees of text coverage that may be provided if learners are able to deal with different levels of the scheme (i.e., the text coverage that may be provided if learners can deal with Level 2 affixes only, with Level 3 affixes also, and so on).

It should be noted that despite the fact that, as reported earlier, the validity of the levels in the word families scheme has been questioned, in the absence of any comprehensive data on affix difficulty, the scheme's levels were made use of in this analysis.

3 Results

3.1 Size and Complexity of the Lists

Table 2 gives the mean number of word forms that are included in each band. The number of word forms per family differs significantly across the

Table 2. Descriptive Statistics for the Size of Word Families

	Mean number of word forms	SD	Minimum	Maximum
1K	6.35	4.343	1	35
2K	5.59	3.536	1	22
3K	4.52	2.761	1	18
4K	4.29	2.700	1	19
5K	3.99	2.600	1	23

bands ($H(4) = 277.48$, $p < 0.001$, $\eta^2 = 0.05$), with the 1K families containing the most members and the number decreasing across subsequent bands. It can also be seen that some word families are very large, the largest, ORGANIZE, having 35 word forms, while some consist of just a single member (e.g., ABOUT in the 1K band, ABOVE in the 2K band, ABROAD in the 3K band, ALIKE in the 4K band, and ABOARD in the 5K band). The 1K band, despite having the highest mean number of members, contains 104 families consisting of a single member, approximately twice as many as the other bands (2K = 49; 3K = 50; 4K = 52; 5K = 51). This is due to the large number of function words in the 1K band. Accordingly, the multi-member word families in the 1K band contain an average of almost seven members per family, while those in the 5K band have just over four members.

Table 3 shows the number of word forms in each band containing different numbers of affixes. That is, there are some forms that feature one affix (e.g., *educat-ion*, in EDUCATE, a 1K family), some with two affixes (e.g., *educat-ion-al*), some with three affixes (e.g., *educat-ion-al-ist*), and some with four affixes (e.g., *educat-ion-al-ist-s*). A Pearson's chi-square shows a significant difference between the bands in this regard ($\chi^2(16) = 602.09$, $p < 0.001$, Cramer's $V = 0.09$, a small to medium effect), and examining the standardized residuals (which reveal the difference between the observed value in each cell of the table and the value that can be predicted on the basis of the overall figures) allows the location of these differences to be pinpointed. This first reveals that the considerable number of forms with zero affixes (irregular verb and noun forms, abbreviated forms, alternative spellings, and compound forms, as mentioned above) do not occur evenly across the bands. There are significantly more in the 1K band, primarily due to the presence of irregular verb and noun forms. Second, while in all five bands the majority of forms contain a single affix, there are differences across the bands. Specifically, there are significantly fewer forms with a single affix in the 1K band and significantly more such forms in bands 3K–5K. Corresponding with this, there are significantly more forms with two and three affixes in the 1K and 2K bands and significantly fewer such forms in bands 3K–5K. Finally, it may be pointed out that the number of forms with two or more affixes (5307) is not trivial, accounting for 26.9% of the forms overall, and that there are 1877 forms containing two or more derivational affixes.

Table 4 presents the number of forms within each band across different levels of the Bauer and Nation scheme. For example, the 1K band has 3040 word forms that feature only Level 2 affixes (e.g., *admitt-ed* and *clean-ing*), a further 1097 forms with Level 3 affixes only (e.g., *un-clean*) or Levels 2 and 3 affixes

Table 3. Number of Forms with Various Numbers of Affixes

	Forms with zero affixes (excluding base words)	Forms with one affix	Forms with two affixes	Forms with three affixes	Forms with four affixes
1K	433 (8.1)	3223 (60.3)	1495 (28.0)	188 (3.5)	9 (0.2)
2K	122 (2.7)	3068 (66.8)	1257 (27.4)	135 (2.9)	11 (0.2)
3K	99 (2.8)	2613 (74.3)	751 (21.4)	53 (1.5)	1 (0.0)
4K	72 (2.2)	2451 (74.5)	717 (21.9)	47 (1.4)	1 (0.0)
5K	65 (2.2)	2284 (76.3)	603 (20.2)	39 (1.3)	0 (0.0)
1–5K	791 (4.0)	13639 (69.1)	4823 (24.4)	462 (2.3)	22 (0.1)

Note: Brackets show the proportion of forms in each band (excluding base words) with each number of affixes.

Table 4. Number of Forms at Different Levels of Bauer and Nation's Scheme

	Forms with Level 2 affixes only	Additional forms through to Level 3 affixes	Additional forms through to Level 4 affixes	Additional forms through to Level 5 affixes	Additional forms through to Level 6 affixes ^a
1K	3040 (56.8)	1097 (20.5)	373 (7.0)	218 (4.1)	187 (3.5)
2K	2921 (63.6)	808 (17.6)	388 (8.4)	163 (3.5)	191 (4.2)
3K	2431 (69.1)	552 (15.7)	236 (6.7)	95 (2.7)	104 (3.0)
4K	2315 (70.4)	444 (13.5)	217 (6.6)	112 (3.4)	128 (3.9)
5K	2104 (70.3)	396 (13.2)	226 (7.6)	95 (3.2)	105 (3.5)
1–5K	12811 (64.9)	3297 (16.7)	1440 (7.3)	683 (3.5)	715 (3.6)

Note: Brackets show the proportion of forms at each band (excluding base words).

^aForms in the lists featuring affixes outside Bauer and Nation's scheme (see *Other affixes observed* column in the additional material online) are included here with Level 6 affixes.

(e.g., *admitt-ed-ly*), and so on. It should be noted that this analysis is based purely on the level of the affixes in the scheme, not on the number of affixes. Thus, for example, among the 3040 forms in the 1K band with Level 2 affixes only, there are 72 forms that feature two Level 2 affixes, such as *find-ing-s* and *low-er-ed*.

A Pearson's chi-square on the Table 4 figures finds a significant difference across the bands in the number of forms with different levels of affixes: $\chi^2(16) = 210.08$, $p < 0.001$, Cramer's $V = 0.05$, a small effect. The standardized residuals reveal that, in general, a similar proportion of the forms in each band are accounted for by the affixes through to Level 4, Level 5, and Level 6. Where the bands differ is in the proportion of forms with Level 2 affixes only and the proportion with affixes through to Level 3. The 1K band contains significantly fewer forms with Level 2 affixes only in comparison with the overall trend across the five bands, while the 3K–5K bands contain significantly more. As for forms with affixes through to Level 3, the 1K band contains significantly more such forms and the 4K and 5K bands significantly fewer.

A final aspect of the complexity of the word families concerns the number of families for which the base word is the most frequent member of the family and the number for which it is not. For example, in the BNC the word form *boy* is the most frequently occurring of the four forms in the BOY word family. There are

Table 5. Estimates of the Number of Families for Which the Base Word is the Most Frequent Member

	Base word is the most frequent member	Base word is not the most frequent member
1K	820	180
2K	740	260
3K	810	190
4K	760	240
5K	750	250
1–5K	3880	1120

Note: Estimates are based on samples of 100 word families from each band.

20 807 occurrences of the four forms: 12 714 (61.1%) for *boy*, 7790 (37.4%) for *boys*, 159 (0.8%) for *boyish*, and 144 (0.7%) for *boyhood*. In the ACTIVE word family, in contrast, the most frequently occurring of its nine forms is *activities*, with 11 476 (34.3%) out of 33 469 total occurrences for the family, while the form *active* itself has 7219 (21.6%) occurrences. As Table 5 shows, on the basis of samples of 100 word families from each band, for the majority of word families the base word is the most frequently occurring member. Nonetheless, in over 20% of the families, and fairly consistently across all five bands, another form occurs more frequently.

3.2 Coverage Levels

On the coverage levels that may be provided by the lists, Table 6 gives the mean proportion of the total occurrences in the BNC of all the forms in a word family that are accounted for by occurrences of the base word alone. For example, as explained above, in the BNC the base words of the BOY and ACTIVE word families account for 61.1% and 21.6%, respectively, of the total occurrences of all the forms in their families. As the table shows, on average the base words account for around three fifths of the occurrences of each family, with similar figures across all five bands. The standard deviations and minimum and maximum figures also show, however, that there is a great deal of variation among the families. There are families for which the base word accounts for 100% of occurrences, this obviously being the case for families consisting of a single member, and families for which the base word accounts for a very small proportion of the occurrences, the lowest being 0.8% for the GOVERN family.

Table 7 shows the proportion of occurrences accounted for by forms up to and including each level of Bauer and Nation's scheme. Note that other forms, that is, irregular verb and noun forms, abbreviated forms, and so on, have been included prior to any of the levels of affixation, though it is unclear if this is appropriate or not. As can be seen, the addition of the Level 2 forms results in a large increase in the proportion of occurrences accounted for, with smaller increases with the addition of each subsequent level.

Table 8 illustrates how the above figures can be used to revise coverage estimates. For example, if the word lists provide 95% coverage of a given text, but

Table 6. Estimates of the Proportion of the Occurrences of Families Accounted for by the Base Word only

	Mean	SD	Minimum	Maximum
1K	67.0	27.52	0.8	100
2K	56.7	26.22	0.9	100
3K	63.0	27.48	1.1	100
4K	62.9	28.53	2.1	100
5K	59.1	27.83	2.2	100
1–5K	62.0	27.46	0.8	100

Note: Estimates are based on samples of 100 word families from each band.

Table 7. Estimates of the Mean Proportions of the Occurrences of Families Accounted for by Forms at Different Word Family Levels

	Base word only	Plus other forms ^a	Plus Level 2 forms	Plus Level 3 forms	Plus Level 4 forms	Plus Level 5 forms	Plus Level 6 forms
1K	67.0	71.2	87.9	91.7	95.6	97.2	100.0
2K	56.7	57.5	82.5	89.4	92.7	95.1	100.0
3K	63.0	64.0	87.8	93.1	95.6	96.5	100.0
4K	62.9	63.1	86.6	91.0	93.1	94.7	100.0
5K	59.1	60.5	88.2	90.4	94.3	96.7	100.0
1–5K	62.0	63.2	86.6	91.1	94.3	96.0	100.0

Note: Estimates are based on samples of 100 word families from each band.

^aIrregular verb and noun forms, abbreviated forms, alternative spellings, and compound forms.

Table 8. Estimates of How Assumed Coverage Levels are Affected by Different Levels of Affix Knowledge

Assumed coverage	If base words only are known	Plus other forms	Plus Level 2 forms	Plus Level 3 forms	Plus Level 4 forms	Plus Level 5 forms	Plus Level 6 forms
95	58.9	60.1	82.3	86.6	89.5	91.2	95.0
98	60.8	62.0	84.9	89.3	92.4	94.1	98.0

learners are only able to deal with Level 2 of the scheme, the actual coverage is estimated to be 82.3%. This is calculated by multiplying the assumed coverage figure (i.e., 95%) by the mean proportion of the occurrences accounted for by including up to Level 2 forms from Table 7 (i.e., 86.6).

As can be seen, if learners are unable to deal with the challenges of WF6, there is a considerable impact on the actual levels of coverage that may pertain. Even if learners are able to cope with affixes at Level 5, but not at Level 6, the impact is substantial: a text assumed to have 95% coverage (one in twenty words unknown) may in fact have only 91% coverage (almost one in ten words unknown), and a text assumed to have 98% coverage (one in fifty words unknown) may have only 94% coverage (three in fifty words unknown).

4 Discussion

The first question this study sought to answer was: To what extent do WF6-based word lists have characteristics that may be challenging for learners? The results above provide a number of insights into these characteristics.

It was found that the 5000 word families in the BNC-based lists contain almost 25 000 word forms, with the 1K band alone containing over 6000 forms despite 10% of its families being single-member families. Among these word forms, there are a number with zero affixes, consisting of irregular verb and noun forms, abbreviated forms, alternative spellings, and compound forms, which it is presumed are included on the assumption that they cause no problems for learners. There appears to be no research on whether this is actually the case or not. However, it must be said that while these forms may appear transparent, learners may have a different perspective.

Looking at the number of affixes in each word form, it was found that the majority contain just a single affix. Nevertheless, over 5000 forms, 26.9% of the total, contain multiple affixes, and there are almost 2000 forms with multiple derivational affixes. The extent to which learners are able to deal with such forms is an unexplored question. However, it seems likely that the more affixes a form contains the greater the difficulty for learners, since each additional affix makes the link between the base form and the derived form more obscure, both in terms of orthography/phonology and semantics. Some evidence of this was found by S. McLean in his 2017 study. He reports (personal communication, April 17th, 2018) that around a third of learners who demonstrated knowledge of the forms *use*, *reuse*, and *usable* could not do likewise for *reusable*. Also notable in this study is that the 1K and 2K bands contain significantly more forms with two and three affixes as compared with the 3K–5K bands.

Next, the results presented the number of forms accounted for by different levels of Bauer and Nation's word families scheme. This first revealed that among the forms that feature only Level 2 affixes, which are inflectional affixes, there are some containing two affixes (e.g., *find-ing-s* and *low-er-ed*). Thus, forms with inflectional (Level 2) affixes are not necessarily inflections (i.e., while *findings* and *lowered* include inflectional affixes, *findings* is not an inflection of *find* and *lowered* is not an inflection of *low*). This demonstrates that word families are a formal categorization, not a functional or grammatical categorization. Level 2 word families are, therefore, not lemmas nor flemmas, but a somewhat different unit.

Overall, around two-thirds of the word forms feature Level 2 affixes only. Nevertheless, there were significant differences across the bands, with only somewhat over half of the 1K forms featuring Level 2 affixes only, while around 70% do so in the 3K–5K bands.

There are therefore three findings that suggest the higher bands, and particularly the 1K band, may be especially challenging for learners. First, there is a significant difference across the bands in the number of forms. There are simply more forms in the 1K band. Second, the 1K band contains significantly fewer forms with a single affix (e.g., *accept-able*) and significantly more with two and three affixes (e.g., *accept-ab-ly* and *un-accept-ab-ly*). It is likely that such complex

forms are more challenging for learners. Third, the 1K band contains significantly fewer forms with Level 2 affixes only (e.g., *accept-ed*) and significantly more with affixes through to Level 3 (e.g., *un-accept-able*; *un-* and *-able* being Level 3 affixes). The actual difficulty of the affixes in Bauer and Nation's scheme has been questioned (Danilović, Savić, & Dimitrijević, 2013; Mochizuki & Aizawa, 2000), but it is the case that the 1K band has proportionally fewer forms featuring affixes that the scheme itself regards as easier. The 1K band, then, which should presumably be the starting point for learners and might be expected to contain the easiest words, in fact may pose the greatest challenge for learners. This may not be a fault with the lists as such, but more a reflection of the nature of language. It does, however, prompt questions about the appropriateness of having fixed bands of 1000 word families (see Brown, 2017; Kremmel, 2016).

Finally in this section, it was seen that in around one-fifth of the word families, the base word is not the most frequent member of the family. Indeed, in around 10% of the families, a form other than the base form is more than twice as frequent as the base form itself. Notwithstanding the fact that in classroom-based learning the base word may be first encountered despite being of lower frequency, this means that for some word families a derived form is likely to be acquired first. Bauer and Nation (1993) suggest this is unproblematic: "once the base word or *even a derived word* [emphasis added] is known, the recognition of other members of the family requires little or no extra effort" (p. 253). Taking an earlier example, this suggests that if a learner knows *activities* (a derived form), little effort is required to understand *active* (the base form) or *actively* (another derived form).

It should be recalled, however, that Ward and Chuenjundaeng (2009) found that Thai learners who could give a translation of a derived form often could not translate the base form. Moreover, it is unclear how the process of dealing with other derived forms is envisaged. Is the assumption that learners are able to understand *actively* simply on the basis of their knowledge of *activities*? Are learners assumed to reverse-derive so-to-speak *active* from *activities* and then derive *actively* from *active*? Whatever the process, it seems likely that such cases do place an additional burden on learners. In an L1 study, Carlisle and Fleming (2003) discovered that children find it more difficult to recognize a base word within a derived form when the affix is unfamiliar (e.g., the unfamiliarity of the affix *-let* caused difficulty in recognizing the base word *tree* in the form *treelet*). By analogy, it may be that L2 learners find it difficult to recognize an affix within a derived form when the base word in that form is unfamiliar (i.e., a learner may struggle to recognize the affixes *-ity* and *-es* within *activities* if unfamiliar with the base word *active*). Learners may, then, not see a derived form as containing affixes at all. Thus, the 22% of word families in which the base word is not the most frequent form, and therefore less likely to be acquired first, may present learners with additional difficulties.

The second research question asked in this study was: If the assumption that learners can deal with WF6 is incorrect, what levels of text coverage might such lists provide? This was pursued by looking at the proportion of the occurrences of a word family that is accounted for by word forms at different levels of the word families scheme. This analysis was based on BNC data and thus is dependent on the

BNC being broadly representative of texts that learners may encounter. This is also of course true of the word lists themselves since they too were based on the BNC.

It was found that on average the base word alone accounts for 62% of the occurrences of all the members in a word family, Level 2 word forms account for around 23% of occurrences, with subsequent levels accounting for 2–5% of occurrences. These figures allow calculations to be made of the coverage levels that may actually ensue if the assumption that learners can easily deal with WF6 is mistaken. For example, if the word lists provide 95% coverage of a text, but learners are unable to cope with WF6, the actual coverage level may be substantially lower, from 59% if the base words alone are known, to 82% if other forms and Level 2 affixes can be dealt with, to 91% if Level 5 affixes are manageable. The point here is not that a narrower definition of the lexical unit would necessarily lead to lower coverage levels being achieved by a word list of a given size (see the coverage levels achieved by Brezina and Gablasova's (2015) lemma-based New General Service List). Rather the point is that if we assume that WF6 are suitable for learners, but it turns out that they are not, we are overestimating coverage to a degree that means learners are likely to experience far more difficulty than we imagine.

These findings have important implications for any work based on the word family. This includes research on vocabulary coverage and comprehension, vocabulary size requirements for a variety of tasks, incidental learning from reading, the balance between explicit and implicit vocabulary learning, estimating learners' vocabulary size, and setting vocabulary learning goals. For instance, a number of studies have explored coverage levels provided by word-family-based word lists and have attempted to estimate the vocabulary size needed to comprehend different types of text (e.g., Adolphs & Schmitt, 2003; Hsu, 2011; Nation, 2006b; Webb & Macalister, 2013). If learners cannot deal with word families, however, the coverage provided by a vocabulary of a given size may be substantially overestimated and the true vocabulary size needed may be far greater. Another example, as Kremmel (2016) and McLean (2017) have discussed, is the considerable number of vocabulary tests that use the word family as the lexical unit. Many tests assume that if a learner displays knowledge of one member of a word family (often the base form), they also have knowledge of all the other members. If this assumption is mistaken, such tests are considerably overstating the extent of learners' vocabulary knowledge. In short, findings and recommendations based on the word family require reevaluation with respect to whether the assumption that the word family is the most appropriate lexical unit is warranted.

5 Conclusion

This study has considered the choice of the word family as the lexical unit by investigating the characteristics of a set of word-family-based word lists. The lists predominantly contain relatively simple forms featuring a single affix from Level 2 of the Bauer and Nation scheme, but also include a large number of more challenging forms. There are many forms that feature multiple affixes or affixes from the higher levels of the scheme, and in many word families, the base word is not the most frequently occurring member of the family. It was also shown that

if the assumption that learners can deal with the complexity of word families is incorrect, even to a relatively small degree, the apparent coverage levels the lists provide could be quite mistaken.

By laying out the characteristics of word-family-based word lists, this study provides teachers and researchers with a clearer view of the challenge such lists pose and the consequences if learners are unable to meet this challenge. The study's findings raise questions about research and recommendations based on the use of the word family which do not demonstrate that the learners concerned are able to deal with the scale of the challenge posed. This study also provides, in Table 7, a means of estimating the actual coverage levels that may be reached by the word lists for learners that are able to deal with different levels of the Bauer and Nation scheme. Finally, this study suggests that the complexity of the word family means that in many circumstances, alternative units, such as the lemma or flemma, may be more appropriate, even for learners of rather high proficiency (McLean, 2017).

In addition, the study highlights several matters about which information is lacking and which are crucial in attempts to determine the most appropriate lexical unit for learners. These include: (1) whether learners are able to deal with the various types of 'other forms' the lists contain (irregular verb and noun forms, abbreviated forms, alternative spellings, and compound forms); (2) whether learners are able to cope with an unfamiliar derived form when they are only familiar with a different derived form from that family rather than the base word; and (3) whether forms that contain more than one affix pose a particular challenge for learners.

Research on the above matters should provide direction and a more solid basis for the development of word lists. In addition, this study has revealed one other feature which future developers of word lists may consider: the fact that the 1K band appeared more challenging than the other bands in that it contains the most word forms, more forms with multiple affixes, and more forms with what the scheme considers to be more challenging affixes. It may be beneficial if future lists can find some means of making the highest frequency bands, the likely starting point for learners, less rather than more challenging than later frequency bands.

Acknowledgments

The author would like to thank Geoffrey G. Pinchbeck, Brandon Kramer, and Stuart McLean for their comments on an earlier draft of this article.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425–438. doi: 10.1093/applin/24.4.425
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279. doi: 10.1093/ijl/6.4.253
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101–118. doi: 10.1177/0265532209340194

- Bertram, R., Laine, M., & Virkkala, M. M. (2000). The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology*, 41(4), 287–296. doi: 10.1111/1467-9450.00201
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22. doi: 10.1093/applin/amt018
- Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System*, 41(4), 1043–1055. doi: 10.1016/j.system.2013.10.013
- Brown, D. (2017). Coverage-based frequency bands: A proposal. *Vocabulary Learning and Instruction*, 6(2), 52–60. doi: 10.7820/vli.v06.2.Brown
- Carlisle, J. F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading*, 7(3), 239–253. doi: 10.1207/s1532799xssr0703_3
- Cobb, T. (no date). Web Vocabprofile, an adaptation of Heatley and Nation's (1994) *Range*. Retrieved from <http://www.lex tutor.ca/vp/>
- Coniam, D. (1999). An investigation into the use of word frequency lists in computing vocabulary profiles. *Hong Kong Journal of Applied Linguistics*, 4(1), 103–124.
- Danilović, J., Savić, J. D., & Dimitrijević, M. (2013). Affix acquisition order in Serbian EFL learners. *Romanian Journal of English Studies*, 10(1), 77–88. doi: 10.2478/rjes-2013-0006
- Davies, M. (2004). *BYU-BNC: The British National Corpus*. Retrieved from <http://corpus.byu.edu/bnc>
- Gardner, D. (2007). Validating the construct of *word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241–265. doi: 10.1093/applin/amm010
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. doi: 10.1093/applin/amt015
- Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes*, 30(4), 247–257. doi: 10.1016/j.esp.2011.04.005
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. doi: 10.1002/tesq.329
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- McLean, S. (2017). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*. Advance online publication. doi: 10.1093/applin/amw050

- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291–304. doi: 10.1016/S0346-251X(00)00013-0
- Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24(3), 262–282.
- Nation, I. S. P. (2006a). *BNC-based word lists*. Wellington: Victoria University of Wellington. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. (2006b). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La revue canadienne des langues vivantes*, 63(1), 59–81. doi: 10.3138/cmlr.63.1.59
- Nation, I. S. P., & Heatley, A. (2002). *Range: A program for the analysis of vocabulary in texts*. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, P. (2015). Which words do you need? In J. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 568–581). Oxford: Oxford University Press.
- Pinchbeck, G. (2014, March). *Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English*. Paper presented at the American Association of Applied Linguistics, Portland, OR.
- Reynolds, B. L. (2015). The effects of word form variation and frequency on second language incidental vocabulary acquisition through reading. *Applied Linguistics Review*, 6(4), 467–497. doi: 10.1515/applirev-2015-0021
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. doi: 10.1111/j.1540-4781.2011.01146.x
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461–469. doi: 10.1016/j.system.2009.01.004
- Webb, S., & Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, 47(2), 300–322. doi: 10.1002/tesq.70
- Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, 44(3), 263–277. doi: 10.1177/0033688213500582