Article

Researching Vocabulary in the EFL Context: A Commentary on Four Studies for JALT Vocabulary SIG

Stuart Webb
University of Western Ontario
doi: http://dx.doi.org/10.7820/vli.v04.1.webb

Abstract

Four papers by Anna C-S. Chang, Yuko Hoshino, Tatsuya Nakata and Andrew Gallacher were presented in the afternoon session of the 4th Annual JALT Vocabulary SIG Vocabulary Symposium in Fukuoka, Japan on June 20, 2015. As discussant, it is my pleasure to comment upon each manuscript. The four studies investigate different issues related to vocabulary learning: learning vocabulary through extensive reading, measuring vocabulary knowledge, perceptions of vocabulary learning through flashcards, and learning vocabulary through writing. After commenting on each paper in turn, I shall present a few suggestions for future research on each of these important issues.

1 Introduction

There has been a great deal of research on vocabulary in Japan in recent years. Perhaps the reason for this is the close relationship that two leading scholars, Paul Nation and Paul Meara, have had with Japan. Paul Nation has spent a considerable amount of time in Japan based in part on his work at Temple University Japan. His frequent presentations around Japan have influenced many researchers and fueled much of the research on lexis in the Japanese EFL context. Paul Meara's work with the Swansea University MA and Ph.D. programs that focused on lexical studies attracted and established a large number of researchers based in Japan. That program has since moved to Cardiff University under Tess Fitzpatrick's guidance but continues to ignite a passion for vocabulary research and a stream of publications by students based in Japan. It should be of no surprise then that there is a lot of useful research currently going on in the Japanese EFL context, and the studies presented in this issue are an indication of this. Three of the four were conducted in Japan, and the fourth was completed in Taiwan.

The four studies investigate different issues related to vocabulary learning: learning vocabulary through extensive reading (Chang), measuring vocabulary knowledge (Hoshino), perceptions of vocabulary learning through flashcards (Nakata), and learning vocabulary through writing (Gallacher). Together they do a great job of revealing the depth and breadth of research on vocabulary that is being conducted today.

2 The Four Studies

2.1   An Investigation of Different Text Levels on L2 Learners' Vocabulary Learning Rates in an Extensive Reading Program, by Anna C-S. Chang

Although there has been a large amount of research investigating incidental vocabulary learning through reading a single text, there have been relatively few studies of vocabulary learning through extensive reading. The importance of vocabulary growth through reading is well established, and further studies of vocabulary learning through extensive reading are clearly warranted. The challenge of conducting studies of vocabulary learning through extensive reading is that they need to involve participants reading multiple texts over a relatively long period of time. Variables that are typically controlled for in short-term studies may then come into play. Because there are so many highly controlled studies of incidental vocabulary learning through reading a single text, there is strong support for ecologically valid studies such as this one by Anna Chang.

This paper builds on two studies that Anna Chang and I recently completed that look at how prior word knowledge (Webb & Chang, 2015) and frequency and distribution of occurrence (Webb & Chang, 2014) affect learning. It looks at text level, which is another variable that was examined in the former study, and frequency, which was examined in the latter. Because Webb and Chang's (2014) study showed an absence of a frequency effect in contrast to many earlier studies (e.g., Horst, Cobb, & Meara, 1998; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007), further research here is useful. The paucity of research on text level makes it another useful factor to examine.

There are many interesting aspects of the study. First, although the Level 1 and Level 3 readers are expected to vary considerably in the vocabulary found in the texts, the lexical profiles of the readers are remarkably similar. The primary difference between the proportions of items in the different frequency bands in the two levels is that there are a greater proportion of proper nouns in the Level 1 texts than in the Level 3 texts. In fact, the proportion of words at the 2000 word level (5.61%) is exactly the same in the two sets of five graded readers. The similarity between the two sets of graded readers is somewhat surprising but supported by the findings of an earlier study that I conducted with John Macalister. Webb and Macalister (2013) found that there was little difference between the lexical profiles of Stage 1, 2, 3, and 4 graded readers in the Oxford Bookworms series.

The use of audio support during the extensive reading treatment also deserves mention. Research indicates that vocabulary learning gains are likely to be greater for reading while listening than reading alone (Brown, Waring, & Donkaewbua, 2008; Webb & Chang, 2012) suggesting that audio support should be included whenever possible. Audio support may account to some extent for the extremely high gains in this study in comparison to those of earlier reading studies (e.g., Horst et al., 1998; Pigada & Schmitt, 2006; Waring & Takaki, 2003).

One aspect of the methodology that could use more detail is the tests for each block of target words. The blocks included 8–15 target words, but was the number

of target words in each block controlled between the tests for Level 1 and Level 3 items? If block size was smaller for words in the Level 1 readers that may make them slightly easier than for the Level 3 readers.

It is generally useful to use multiple tests to measure vocabulary learning. In this study, two tests were administered to participants. The first test is a contextualized receptive translation test. Justification for including context in this test format is that it may help to cue recall, making it more sensitive to learning gains. This is usually a good thing. A limitation in this study, however, is that the presence of the context may contribute to learning. The results would actually suggest that learning may have occurred through completing the test because the gains were higher on this test than in the matching test that followed. Research indicates that the recall test was likely to be the more difficult of the two (Laufer & Goldstein, 2004). In the example provided in the study (Carol is never late for rehearsals, and she knows that these concerts are important for us.), even if we do not know the target word *rehearsal*, we may learn that it is semantically associated with *concerts*, and that perhaps *late* is a collocate. Completing this test may then affect the validity of the results of the second test. Ideally, the receptive translation test would have been decontextualized to ensure that there was no learning effect from taking it.

It is important to note that vocabulary learning may not have been purely incidental. There was the potential for participants to deliberately learn words in different ways, by looking up entries in dictionaries, by asking classmates, and by using newly encountered words in follow-up activities. It would be unethical to not allow participants to do these things and their use mirrors that of extensive reading programs. Thus, the lack of control may increase ecological validity but may muddy the findings (in terms of incidental vocabulary learning) to some degree.

Another possible factor that may have affected findings is the difference between the treatment times for Level 1 and Level 3 graded readers. In the former, it was 5 weeks, and in the latter it was 8 weeks. The greater interval for Level 3 graded readers may have reduced the amount of learning because knowledge of target words that were encountered in the first weeks of the treatment had greater chance to decay for this set of words.

The lack of a relationship between frequency of occurrence and vocabulary learning supported those of the earlier study by Webb and Chang (2014). However, I wondered whether looking at the data in a different way might indicate a relationship between the two variables. Rather than looking at them for the individual sets of words, perhaps combining the two sets and then looking at gains according to frequency bands (1–5 occurrences, 5–10 occurrences, 11–20 occurrences, and 20+ occurrences) might have provided slightly different results.

The results revealed extremely high relative learning gains. Gains ranged from 79.33% to 93.22% at the end of the treatment and from 71.43% to 81.92% on the delayed posttests. These gains exceed those of earlier studies and are in fact higher than those of many studies of deliberate vocabulary learning. The size of the gains might be supported by the use of audio support, as well as the possibility of dictionary use and the use of words in post reading activities. I wondered if other reasons for such large gains were learning effects from taking the tests, awareness of

the focus on vocabulary learning for the second set of target words, and because this was the second year that the participants were involved in extensive reading, perhaps an awareness of an aim within the extensive reading program to learn vocabulary. Despite these possible limitations, the results provide strong support for the use of extensive reading as a means of increasing L2 vocabulary growth.

I enjoyed reading this paper. There is a real need for longitudinal studies of vocabulary learning through extensive reading. I hope that the paper helps to show some of the issues that need to be considered in this line of research.

2.2   Measuring Knowledge of Words with Multiple Meanings, by Yuko Hoshino

There have been a large number of studies of teaching and learning vocabulary that have been conducted in recent years. However, within this research, one topic that has really lagged behind has been the development and validation of new tests of vocabulary knowledge (Webb & Sasao, 2013). Moreover, when new measures of vocabulary have been created, they have almost exclusively focused on measuring breadth of vocabulary knowledge. The focus on developing a test designed to measure a component of depth of vocabulary knowledge (knowledge of different meanings of words) in this paper is original, interesting, and of great value.

The paper provides useful discussion on the lack of research on the learning of multiple meanings of words. The issue of polysemy has been neglected in most studies of teaching and learning vocabulary. Research has provided little or no discussion of the degree to which peripheral meanings are known. Similarly, tests of vocabulary knowledge focus on measuring whether test takers are able to link L2 forms to their core meanings. The extent to which peripheral meanings are known is not measured, nor is there discussion or acknowledgment that test takers may not know these meanings when they know the core meanings.

Issues with creating a test designed to measure multiple meanings are clearly laid out, demonstrating the challenges of creating such a test. The first two issues, ambiguity resolution and grammatical knowledge, affect the degree to which test takers are able to identify target meanings. A reasonable solution is presented by the inclusion of context within the test items. The presence of context allows the tester to isolate the target meaning. The third issue relates to the distance in meaning between the prototypical meaning and the peripheral meaning. This may be the biggest challenge with creating a test that can be applied to randomly generated words. When considering words with multiple meanings, we typically think of items that have a large distance between them. *Admit*, *rose*, and *duck* are three examples. However, polysemous high-frequency words often have many meanings and often the differences between these meaning are very slight. For example, *go*, *play*, and *run* all have a large number of meanings of which several are very close. The prototypical meaning of run might be to move quickly on foot. However, someone can also *run in a race*, and *he was running for his life* conveys a meaning that is beyond simply moving quickly on foot. Providing context to help signal target meanings may again be the key to isolating target meanings. However, it may be

impossible to measure knowledge of these slight differences in meaning in a single test, because answering a question about one meaning may affect a subsequent response about a different meaning. A possible solution might be to have multiple tests. The final issue that is discussed, cross-language differences, may affect learners with different first language backgrounds. Variation in overlap between target meanings in different languages may make interpretation of test results difficult in ESL contexts where test takers may have different language learning backgrounds.

Three test formats were piloted: selection, translation, and pairing. These formats were examined in two ways, with full sentence and collocation cues. In the selection test, the test takers need to select sentences (or collocations) from a number of options that includes distracters. Two correct options are provided for each of three target meanings (six in total) and two distracters are also provided. In the translation test, test takers are required to translate the collocations into their first language. Again two examples are provided for each target meaning. The pairing test used a matching format with clusters of items. Test takers needed to match two sentences (or collocations) that included the same meanings from six different options.

Although data are not presented, we find that the most demanding of the test formats (translation), where sentences rather than only collocates were included, provided the greatest reliability and discriminability. This result might be expected because the other two formats add some degree of chance to scoring correctly; it may be possible to successfully guess the answers in the selection and pairing formats, but test takers are unlikely to translate successfully.

One question that arises when looking at the tests is why two examples of each target meaning are necessary in the selection and translation formats. It might also be fair to question whether the two examples are perceived as being identical in meaning. For example, is the meaning of *broke* in *broke the surface* and *broke the skin* identical? It might be argued that the former involves something entering, whereas the latter involves something opening. One thing that stood out in the test is that the final item is written in the present tense (break my promise) while all of the other items were presented in the past tense (e.g., broke the law). It would be best to use the same verb tense for all items.

Another question that I had about the test was whether the frequency of the words used in the sentences and collocations was controlled. I wondered whether some words such as *public*, *scene*, *law*, *surface*, and *promise* would be known. I assume that these words were included because they collocated with the target word. However, it does indicate that finding simple contexts is likely to be another challenge with creating a test format that would be appropriate for learners at varying levels.

Perhaps the biggest question about these three formats is to what extent they were measuring receptive knowledge of collocation rather than receptive knowledge of concept and referents? Because the varying meanings are always defined by collocations, it may be that it is knowledge of collocation that has the greatest impact on test performance. The proposed test formats certainly indicate that

collocation may be the key to defining peripheral meanings so perhaps the two components of lexical knowledge are intertwined.

Two more test formats, isolation and translation recognition, are also discussed. In the isolation format, test takers are required to produce as many meanings for words as possible. In essence it is a receptive recall test that involves recalling multiple meanings for each target item. This may be the ideal format as it clearly isolates the intended aspect of knowledge. However, as noted, it is also an extremely difficult format. Would you be able to produce the slight differences in meaning for the word *run* without the aid of context or collocations to cue recall? I believe that most native-speakers of English would have trouble scoring well. In the translation recognition format, test takers must indicate as quickly as possible whether meanings that are presented are correct. The focus on speed suggests that this test may be tapping into fluency or strength of knowledge.

I really enjoyed reading this paper. Measuring productive knowledge of multiple meanings is something that I had considered, but I had not gotten very far with, because there are indeed many challenges. Yuko Hoshino's investigation of this topic is a really useful step toward creating a test that would enhance our ability to measure depth of knowledge. I hope that she will continue on with this research.

2.3   Are Learners Aware of Effective Ways to Learn Second Language Vocabulary from Retrieval? Learners Perceived Effects of Relative Spacing, Absolute Spacing, and Feedback Timing on Vocabulary Learning, by Tatsuya Nakata

This paper follows up a series of carefully controlled studies that have examined how different variables affect vocabulary learning using flashcards (Nakata, 2015a; Nakata, 2015b; Nakata & Webb, 2015). This is a very useful line of research. Studies have consistently shown that learning vocabulary through a paired-associate learning approach contributes to impressive gains that are very time efficient (e.g., Webb, 1962; Webb, 2009a, 2009b). Researchers that have focused on teaching and learning vocabulary appear to overwhelmingly support a broadly inclusive approach that involves learning through both meaning-focused and form-focused conditions (e.g., Hunt & Beglar, 2005; Laufer, 2003; Nation, 2001; Schmitt, 2008; Webb, 2013; Webb & Nation, 2013). However, there may still be a perception among some researchers that decontextualized learning conditions such as flashcard learning do not deserve a place in the language learning curriculum. This may be one reason for a lack of studies in recent years on word card and flash card learning in applied linguistics and TESOL journals.

Learner perceptions of the effects of vocabulary learning methods is a useful area of research. Although there has been an abundance of studies on vocabulary and the effects of different learning conditions, there is relatively little known about which approaches to learning are perceived as being most useful. It may often take a long time for pedagogically useful research findings to reach the classroom, so investigating which activities are perceived as being effective may provide some insight into how students might learn on their own outside of the classroom.

The results of Nakata's study indicate that the Japanese students who took part in the research were able to accurately judge the effectiveness of the different learning variables; both equal and expanding spacing and immediate and delayed feedback were viewed as equally effective, while spaced learning was seen as being more effective than massed learning. These findings contrast earlier studies that have indicated that students may not accurately rate the learning effects of different conditions. Nakata attributes the differences between his findings and those of earlier findings to variation in the participant profiles.

With data limited to a relatively short questionnaire, it may be fair to question whether participants clearly understood the constructs examined, and if they did, whether their responses can be generalized to slightly different conditions. Ideally, follow-up interviews would be conducted with a sample of the participants to better explore their perceptions of the learning conditions.

One question that we might ask about the findings is whether we can generalize the participants' responses beyond the very carefully controlled spacing conditions to different types of spacing. For example, when we think of spaced learning with flashcards (or any other vocabulary learning condition), we may often consider spacing to occur in days or over multiple learning periods. However, spacing in the studies that were examined was relatively short and occurred between encounters with different items during a single learning period. Thus, it may be useful to look further at perceptions of different types of spacing. Similarly, the variable of feedback timing that was examined in the study involved manipulation of feedback within a relatively short interval in a single learning session. Although these were ecologically valid timings for feedback, there are certainly other variations of feedback manipulation both within and between sessions that would be useful to explore.

Another issue that should be considered when investigating student perception of learning techniques is how feedback on performance in the relevant learning conditions might affect responses. In Nakata's study, when the participants completed the survey, they did not know how they performed on the tests. In the classroom, we might expect learners to base their perceptions on the gains that they have made through using certain techniques. The degree to which the participants in this study knew how the different variables affected their learning performance is not clear. The results do indicate that they were making judgments that reflected their learning performance. However, investigating how knowledge of learning performance affects responses would also be useful.

Learning through electronic applications allows researchers extreme precision in the control of variables. The flashcard/paired-associate learning condition lends itself very well to this type of research. However, because it is one of many different vocabulary learning techniques, it would be useful to examine how the variables examined in Nakata's research (relative spacing, absolute spacing, feedback timing, block size) apply to other learning conditions. Examining student perceptions of different vocabulary learning techniques would be a useful follow-up to this research.

I enjoyed reading this study. There is much research still be done on learning with flashcards. This paper reveals another avenue for research in this area.

2.4   *Mastery Sentences: A Window into the Interplay between Word Knowledge Types, by Andrew Gallacher*

Students are often asked to do sentence production exercises, where a target word is given and they have to use that word in writing. These exercises typically result in a wide range of responses. For example, responses for the target word *banana* might include: *It is a banana. I like bananas. Bananas are a yellow fruit* and *Monkeys love to pick bunches of bananas from trees and peel and eat the sweet yellow fruit*. Evaluating responses in sentence production exercises is difficult because the sentences demonstrate different degrees of vocabulary knowledge. In the examples above, the only aspects of knowledge that are demonstrated in the first two sentences are written form and grammatical functions. In the third sentence, greater knowledge is demonstrated as knowledge of aspects of meaning are also revealed, as well as collocation. The final sentence, however, conveys a greater degree of vocabulary knowledge because knowledge of aspects of meaning are clearly conveyed and some strong collocates and associations are also included. It is only this final sentence that Gallacher might classify as a mastery sentence.

The concept of mastery sentences is an interesting one for several reasons. First, providing a concrete goal for learners to move beyond knowledge of form-meaning connection and use words correctly in sentences is a useful learning target. As Gallacher notes, many learners may be unable to use words effectively, particularly in the EFL context where opportunities for meaning-focused output are limited. Second, if mastery sentences are learned together with new target words, then perhaps the context may help to cue recall of the target words increasing retention. Moreover, if the sentence can be recalled, then it may provide the foundation necessary for creative use, helping learners to generate the target word in original sentences. Third, perhaps mastery sentences may be a useful way to reduce the effects of cross-association when words are learned in lexical sets. Paul Nation suggested that when there is no option but to learn semantically related words together, it is best to learn each word in a sentence that helps to differentiate it from other words (Nation, 2000). Mastery sentences aim to use target words in a somewhat unique way, where only direct synonyms might replace the target word. This should help to reduce the effects of interference.

The challenge with mastery sentences may be that they are difficult to generate. If we look at the examples provided in the appendix for the target words *solve* (Police officers should solve all problems) and *curious* (Children are curious about many things), although they do demonstrate productive knowledge of aspects of use, they do not meet the requirement that only direct synonyms can be used to replace the target words. We might replace *solve* with *investigate*, *consider*, *examine*, and *discuss*, whereas we could replace *curious* with *anxious*, *informed*, *told*, and *asked*. This makes me wonder whether it would be best for teachers to provide an initial mastery sentence for each target word that might serve as a model to help students generate their own sentences.

Examining the degree to which the mastery sentences demonstrated knowledge of meaning and usage was interesting. However, scoring the correct usage of target words is challenging. Exemplars that demonstrate the criteria for how to successfully score mastery sentences would be useful. One criterion that is reported

in this paper indicates that providing a correct collocation in the sentence is necessary for usage (peel and yellow were used with the target word banana; eat and soup were used with the target word spoon; occurrence was used together with common). In fact I wondered if it would be possible to create a mastery sentence without a correct collocate because collocations may be necessary to reduce the potential to substitute other words for the target word. However, the inclusion of correct collocates (occurrence used together with common) alone is insufficient because the clause that the target word appears in must be grammatically correct. One question that I had when looking at the examples was whether sentences should be incorrect if a target word is used correctly but if there are errors in the clause relating to other words. It might be argued that the omission of the article *a* before *common occurrence* had greater relevance to knowledge of the collocate *occurrence* than it did for the target word *common*. Another way to score the sentences might be to focus on the aspects of use in Nation's (2001) description of vocabulary knowledge (grammatical functions, collocations, constraints on use) that are demonstrated in a sentence. In the high-level sentences provided in the paper that were scored as incorrect usage, both sentences might be scored correctly for knowledge of grammatical functions. The first would be scored correctly for collocations (common occurrence) and the second would not.

Examining the lexical frequency profile of the sentences was useful. The greater use of lower frequency words by students at higher proficiency levels supports earlier findings (e.g., Laufer & Nation, 1995). The finding that it is the use of these lower frequency words that causes errors is quite an interesting finding. In lexical frequency profiling studies errors are often ignored. This study provides some indication that it would be useful to examine not only the use of lower frequency words in writing but also the degree to which the use of lower frequency words develops. Larger quantitative studies of mastery sentences would have probably hidden this interesting finding.

Although this study is small in scale, the results provide an interesting picture of how two aspects of vocabulary knowledge develop in writing. There is a need for qualitative studies of vocabulary (Nation & Webb, 2011). They may provide a more accurate indication of learning within a program and may also provide insight into vocabulary development that is not revealed in larger quantitative studies.

3 Suggestions for Further Research

Although the four papers discussed in this issue all focus on vocabulary, they each look at very different topics: vocabulary learning through extensive reading, measuring vocabulary knowledge, perception of the effects of vocabulary learning through flashcards, and the use of words in student writing. Each of these areas deserve further investigation. I will briefly touch on how each of these areas might be examined in further research.

There have been many studies of vocabulary learning through reading a single text. Although researchers are typically cautious in generalizing their findings to larger scale reading programs, the findings often become recognized as being typical of extensive reading. This is worrying because the results of the few studies of vocabulary learning through extensive reading have provided somewhat different

findings than those of more carefully controlled short-term studies. There is thus a need for more longitudinal studies of vocabulary learning through extensive reading. One reason for the lack of such studies is that they may be more challenging to conduct. A greater number of variables may come into play in longitudinal research so it is more difficult to attribute findings solely to the treatment. However, if we are to clearly determine the value of extensive reading to vocabulary growth, studies that involve large amounts of reading are necessary.

The development and validation of new tests of vocabulary knowledge are also needed. In recent years (and in this issue), there has been greater focus on the development of depth of vocabulary knowledge. Yet, there have been few tests designed to measure depth as a whole or individual aspects of knowledge besides form-meaning connection. With much greater focus on learning multi-word items and formulaic language, tests designed to measure collocation would be useful. Moreover, it would also be useful to measure students' skill at learning words. Yosuke Sasao's (2013) research in this area was particularly innovative and a useful model for further studies.

Despite consistent findings indicating the value of flashcard learning, learning vocabulary with flashcards has been neglected within the fields of applied linguistics and TESOL. One key question that remains to be answered is how many words can be learned with flashcards over different periods of time? Finding this answer would provide useful support for the inclusion of flashcards within the language learning curriculum. It should be noted that flashcards on their own are not the answer to vocabulary learning. However, they provide a starting point for the development of lexical knowledge. Knowledge of target words may then be expanded through further meaning-focused learning tasks. Another useful question to explore is the degree to which the presence of context within a flashcard might contribute to learning.

There is always a lot of interest in the use of words in writing. In recent years, this has been driven to some degree by new software such as RANGE (Nation & Heatley, 2002), AntWordProfiler (Anthony, 2014), and Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) that allows us to analyze large amounts of written text in great detail. TAALES (Kyle & Crossley, in press) is a new application that has evolved and improved on Coh-Metrix. It includes 135 indices of lexical sophistication and richness and may become the next key tool to research the use of words in writing. However, Gallacher's study in this issue also reveals that smaller scale studies might provide insight that is not apparent in these much larger scale studies. Large scale quantitative studies, and qualitative studies that look at the development of different aspects of vocabulary knowledge in L2 writing are still needed.

References

Anthony, L. (2014). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/

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

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. doi:10.3758/BF03195564

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, *17* (1), 23–59.

Kyle, K., & Crossley, S. A. (2014). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*. doi:10.1002/tesq.194

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence? *Canadian Modern Language Review*, *59*, 567–587. doi:10.3138/cmlr.59.4.567

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. doi:10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, *16*, 307–322. doi:10.1093/applin/16.3.307

Nakata, T. (2015a). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, doi:10.1017/S0272263114000825

Nakata, T. (2015b). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*, *19*, 416–434. doi:10.1177/1362168814541721

Nakata, T., & Webb, S. (2015). Does studying vocabulary in smaller sets increase learning? Effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, doi:10.1017/S0272263115000236

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, *9* (2), 6–10.

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

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*, 1–28.

Sasao, Y. (2013). *Diagnostic Tests of English Vocabulary Learning Proficiency: Guessing from Context and Knowledge of Word Parts*. Unpublished PhD Thesis, Victoria University of Wellington, Wellington, New Zealand.

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

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Webb, W. B. (1962). The effects of prolonged learning on learning. *Journal of Verbal Learning and Verbal Behavior*, *1*, 173–182. doi:10.1016/S0022-5371(62)80026-7

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28* (1), 46–65. doi:10.1093/applin/aml048

Webb, S. A. (2009a). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review*, *65*, 441–470. doi:10.3138/cmlr.65.3.441

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

Webb, S. (2013). Depth of vocabulary knowledge. In C. Chappelle (Ed.), *Encyclopedia of applied linguistics* (pp. 1656–1663). Oxford, UK: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal1325

Webb, S. A., & Chang, A. C.-S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68* (3), 267–290. doi:10.3138/cmlr.1204.1

Webb, S. A., & Chang, A. C.-S. (2014). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*. 1–20. Published online before print November 28, 2014, doi:10.1177/1362168814559800

Webb, S., & Chang, A. C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, doi:10.1017/S0272263114000606

Webb, S., & Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, *47*, 300–322. doi:10.1002/tesq.70

Webb, S., & Nation, I. S. P. (2013). Teaching vocabulary. In C. Chappelle (Ed.), *Encyclopedia of applied linguistics* (pp. 5670–5677). Oxford, UK: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal1177

Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, *44*, 263–277. doi:10.1177/0033688213500582