

2016

Vocabulary Learning Exercises: Evaluating a Selection of Exercises Commonly Featured in Language Learning Materials

Tatsuya Nakata
Kansai University

Stuart Webb
The University of Western Ontario, swebb27@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/edupub>



Part of the [Education Commons](#)

Citation of this paper:

Nakata, T., & Webb, S. A. (2016). Vocabulary learning exercises: Evaluating a selection of exercises commonly featured in language learning materials. In B. Tomlinson (Ed.), *Second Language Acquisition Research and Materials Development for Language Learning* (pp. 123-138). Oxon, UK: Taylor and Francis.

Chapter 8 - Vocabulary Learning Exercises: Evaluating a Selection of Exercises Commonly Featured in Language Learning Materials

Tatsuya Nakata and Stuart Webb

Introduction

This chapter examines common second language (L2) vocabulary learning activities using a framework proposed by Nation (2013a). In particular Nation's first guideline which focuses on efficacy will be examined in detail. Nation and Webb's (2011) Technique Feature Analysis will be used to determine which components of the activities contribute to learning. The chapter aims to gauge the relative efficacy of three vocabulary learning activities: learning from flashcards, cloze exercises, and crossword puzzles, shed some light on their strengths and weaknesses, and show how they might be modified to be made more effective.¹

Nation (2013a) argues that vocabulary teaching activities need to meet five guidelines. The first guideline states that activities need to facilitate vocabulary learning (see below for the discussion on how the effectiveness of vocabulary teaching activities are measured). Second, according to Nation, activities should not require a lot of work on the part of teachers. In other words, if there are two activities that are equally effective, and Activity A requires less work from teachers than Activity B, the former is more desirable in terms of practicality. Nation's third guideline states that activities should provide a balance of the four strands of meaning-focused input, meaning-focused output, language-focused learning, and fluency development. Meaning-focused input refers to activities where the focus is on understanding a message such as extensive reading or listening. In meaning-focused output, there is a focus on conveying a message such as giving a speech or writing a story. Language-focused

learning involves intentional learning of an aspect of language such as explicit grammar instruction or memorization of vocabulary. Fluency development refers to activities such as improving the rate of word recognition, reading, or speaking, where the focus is on increasing the speed at which learners use L2 knowledge (e.g., Nation, 2013b). Because having a balance of the four strands is effective for L2 learning, good vocabulary learning activities should also provide a balance of the four strands. Nation's (2013a) fourth guideline states that activities should be efficient. In other words, if there are two activities that are equally effective, and Activity A requires less time than Activity B, the former is more efficient and thus more desirable. Lastly, Nation (2013a) also argues that activities need to be able to be used many times.

Flashcards

Now let us examine common L2 vocabulary learning activities using Nation's (2013a) framework. The first activity to be analyzed is learning from flashcards in the productive direction. Flashcards (also known as word cards) are a set of cards where the L2 word is written on one side and its meaning, usually a first language (L1) translation, is written on the other. Learning flashcards in the productive direction involves viewing the meaning of a target item and then trying to recall its L2 form. Nation's first guideline is concerned with effectiveness. Although it is not very easy to determine which activities are effective for vocabulary development, the present study will analyze the potential effectiveness of activities using a framework known as Technique Feature Analysis (hereafter TFA; Nation & Webb, 2011).

TFA consists of 18 criteria that have been found to facilitate L2 vocabulary learning based on

previous empirical research. If a vocabulary learning activity meets a certain criterion, 1 point is given for that criterion. The more points an activity receives out of a total score of 18, the more effective it is considered to be for vocabulary learning. Table 8.1 shows the results of our analysis of flashcard learning based on TFA.

Table 8.1 Analysis of flashcard learning using Technique Feature Analysis

	Criteria	Explanation	Flashcard
	Motivation		
1	Vocabulary learning goal	Does the activity have a clear vocabulary learning goal?	1
2	Motivate learning	Is the activity motivating for learners?	1
3	Words selected by learners	Do learners select the target words to be studied?	1
	Noticing		
4	Attention on target words	Does the activity encourage the learners to pay attention to the target words?	1
5	Awareness of new vocabulary learning	Does the activity make learners notice new features of target words?	1
6	Negotiation	Does the activity provide opportunities for negotiation?	0
	Retrieval		
7	Retrieval	Does the activity provide opportunities for retrieval?	1
8	Productive retrieval	Does the activity involve productive retrieval?	1
9	Recall	Does the activity involve recall?	1
10	Multiple retrievals	Does the activity involve multiple retrieval opportunities for each target word?	1
11	Spacing between retrievals	Does the activity introduce spacing between retrieval opportunities?	1
	Generation		
12	Generation	Does the activity promote generation?	0
13	Productive generation	Does the activity involve productive generation?	0
14	High degree of generation	Does the activity involve a high degree of generation?	0
	Retention		
15	Successful form-meaning linking	Does the activity provide opportunities for successful linking of form and meaning?	1
16	Instantiation	Does the activity promote instantiation?	0
17	Imaging	Does the activity promote imaging?	0
18	Avoidance of interference	Does the activity avoid interference between words?	1
	Total score		12

Note. Adapted from Nation and Webb (2011, p. 7)

Let us now consider how many criteria shown in Table 8.1 flashcard learning satisfies. The first three criteria in TFA are concerned with motivation. Criterion 1 asks whether there is a clear vocabulary learning goal in the activity. Because learners use flashcards for the purpose of learning vocabulary, the flashcard activity meets this criterion. The second criterion asks whether the activity is motivating for learners. There may exist conflicting views regarding this criterion. On one hand, some learners tend to perceive rote learning, including flashcard learning, as boring (e.g., Krashen, 1989). On the other hand, some researchers point out that flashcard learning may be motivating because it may give learners a sense of accomplishment (e.g., Mondria & Mondria-de Vries, 1994; Nation & Webb, 2011). The use of computer-based flashcards (e.g., Nakata, 2011, 2013a) may be particularly motivating. Here, following Nation and Webb's analysis, 1 point is given for this criterion. The third criterion asks whether learners select the words to be studied. This criterion is based on the assumption that studying L2 words selected by learners may be more motivating than studying L2 words selected by others such as teachers or materials developers. Unless readymade flashcards are used, learners study L2 words selected by themselves. Hence, 1 point is also given for this criterion. In the case of computer-based flashcards, a wide selection of readymade flashcards are available. Nakata (2011) argues that electronic readymade flashcards are useful because they may allow learners to study many lexical items without the time-consuming task of flashcard creation. Whether the use of computer-based readymade flashcards has a negative effect on learners' motivation is an empirical question that needs to be investigated.

Criteria 4-6 in TFA are concerned with noticing. Criterion 4 asks whether the activity focuses learners' attention on the target words. When learning from flashcards, learners deliberately attempt to learn target words, which involves noticing. Hence, 1 point is given for this

criterion. The next criterion is concerned with whether the activity makes learners notice some features of L2 words that need to be learned. Because flashcard learning typically involves learning (at least partially) unfamiliar L2 word forms and their meanings, flashcard learning also meets this criterion. Criterion 6 is concerned with negotiation. Because flashcard learning is usually done individually, this criterion is not met.

The next five criteria are concerned with retrieval. Retrieval refers to the act of accessing previously learned information about L2 words from memory. Retrieval can be categorized into two types: productive and receptive. The former involves retrieving L2 word forms, whereas the latter involves retrieving the meanings of L2 words. Criterion 7 asks whether the activity involves retrieval. This criterion is based on research showing that retrieval enhances the retention of L2 vocabulary (e.g., Folse, 2006; Barcroft, 2007; Karpicke & Roediger, 2008). Because learning flashcards in the productive direction involves productive retrieval, it satisfies this criterion. Retrieval is also a common feature among existing computer-based flashcards (Nakata, 2011).

The next criterion is concerned with the direction of retrieval (i.e., receptive or productive). Research suggests that if only one direction has to be chosen, productive retrieval may be more desirable than receptive retrieval. This is because productive retrieval results in adequate gains in receptive knowledge as well as large gains in productive knowledge, whereas receptive retrieval results in large gains in receptive knowledge but only small gains in productive knowledge (e.g., Mondria & Wiersma, 2004; Steinel, Hulstijn, & Steinel, 2007; Webb, 2009). Because productive retrieval may be more desirable than receptive retrieval, Criterion 8 gives a point for productive retrieval. In Nation and Webb's (2011) analysis, no point is given to flashcard learning regarding this criterion because they limit their analysis to

receptive flashcard learning. However, as Nation and Webb acknowledge, flashcards can be used to practice productive retrieval if learners look at the meanings and attempt to retrieve L2 words. Furthermore, most computer-based flashcards support not only receptive but also productive retrieval (Nakata, 2011). One point, therefore, is given for Criterion 8 in our analysis.

The next criterion is concerned with the recall / recognition distinction. Recall requires learners to produce L2 word forms or their meanings, whereas recognition asks learners to choose L2 word forms or their meanings from a number of options as in a multiple-choice question. Memory research shows that recall may enhance the retention of L1 vocabulary, reading materials, and lecture materials more than recognition (e.g., Carpenter & DeLosh, 2006; Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007). As a result, TFA considers recall to be a positive feature. Because learning from paper-based flashcards usually involves recall, flashcard learning meets this criterion. In the case of computer-based flashcards, recognition formats are also common (Nakata, 2011). It should be noted, however, that although recall is often considered to be more effective than recognition, L2 vocabulary research has found little or no difference between the effects of recall and recognition (Van Bussel, 1994; Nakata, 2013b). The use of recognition formats among computer-based flashcards, therefore, may not necessarily be a negative feature.

Criterion 10 in TFA asks whether the activity involves multiple retrievals of each target word. By going through a stack of flashcards multiple times, learners can practice repeated retrieval. As a result, flashcard learning satisfies this criterion. This criterion is based on the assumption that vocabulary learning increases as a function of retrieval frequency. Yet,

empirical studies have produced mixed results regarding this assumption. Non-L2 vocabulary research has shown that although multiple retrievals may facilitate short-term memory, they may not necessarily enhance long-term retention (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005). L2 vocabulary research, however, demonstrated the advantage of multiple retrievals 4 weeks after the treatment on productive and receptive recall posttests (Nakata, 2013b). It may be reasonable, therefore, to assume that for L2 vocabulary learning, the benefits of multiple retrievals may persist at least 4 weeks after learning.

Criterion 11 asks whether there is any spacing between retrievals. This is based on the spacing effect. According to the spacing effect, spaced learning, which introduces spacing between retrievals of a certain item, yields superior retention to massed learning, which does not involve any spacing (e.g., Karpicke & Bauernschmidt, 2011; Nakata, accepted). Spacing is found to have a very large effect on learning. Nakata (accepted), for instance, found that spaced learning was more than twice as effective as massed learning on a posttest conducted 1 week after the treatment. Flashcards allow learners to introduce spacing between retrievals unless learners practice the same word repeatedly without any interval. Hence, flashcard learning also satisfies Criterion 11.

Given that introducing spacing increases learning, one might ask how we should space retrieval opportunities in order to maximize learning. Previous research shows that larger spacing generally leads to better long-term retention than shorter spacing (e.g., Pashler, Zarow, & Triplett, 2003; Pyc & Rawson, 2007; Karpicke & Bauernschmidt, 2011; Nakata, 2013b). In other words, studying a set of L2 words every month leads to better long-term retention than studying them every week. This is a phenomenon known as the lag effect. The

lag effect suggests that not only the presence or absence of spacing but also the amount of spacing may affect vocabulary learning.

Some researchers argue that gradually increasing spacing between retrievals (e.g., 1 month, 2 months, 3 months) may maximize L2 vocabulary learning (e.g., Ellis, 1995; Schmitt, 2000; Hulstijn, 2001). This type of spacing is known as expanding spacing. In contrast, a schedule where spacing between retrievals of a given item is held constant (e.g., 2 months, 2 months, 2 months) is referred to as equal spacing. Although expanding spacing is often regarded as the most effective type of spacing schedule, L2 vocabulary studies comparing equal and expanding spacing have produced inconsistent results. Three studies failed to find any advantage of expanding over equal spacing in their posttest scores (Pyc & Rawson, 2007; Karpicke & Bauernschmidt, 2011; Kang, Lindsey, Mozer, & Pashler, in press). Nakata (accepted) showed statistically significant, yet only limited, advantage of expanding over equal spacing. Overall, research suggests that the amount of spacing may have a larger effect on learning than the type of spacing (i.e., expanding or equal; Karpicke & Bauernschmidt, 2011). Because flashcards offer flexibility in the order of items, they may allow learners to manipulate the amount and type of spacing relatively easily, which is another potentially positive feature of flashcard learning. Computer-based flashcards are also useful because they can be programmed to keep a record of a learner's performance and ensure that L2 words are practiced at regular intervals (e.g., Nakata, 2008, 2011).

The next three criteria in TFA are concerned with generation (also referred to as *generative* or *creative use*), where learners meet or use familiar words in novel contexts (e.g., Joe, 1998). In flashcard learning, L2 words are always met in the same context, unless learners make

multiple cards for one word, each of which illustrates its different usages. As a result, flashcard learning meets none of Criteria 12-14. Although computer-based flashcards could be programmed to promote generation, it is not a common feature among existing vocabulary learning programs (Nakata, 2011).

The last four criteria in TFA are concerned with factors affecting retention. Criterion 15 asks whether learners have a chance to be exposed to correct L2 word forms and their meanings. In flashcard learning, learners can verify the correct response by looking at the other side of cards. Flashcard learning, therefore, meets this criterion. Looking at the correct response after retrieval can be considered as a form of feedback. Non-L2 vocabulary research has shown that receiving feedback after a delay (delayed feedback) may increase retention more than receiving feedback immediately after retrieval (immediate feedback), a phenomenon known as the delay retention effect (e.g., Kulik & Kulik, 1988; Butler, Karpicke, & Roediger, 2007; Metcalfe, Kornell, & Finn, 2009). Nakata (in press), however, failed to observe this effect, suggesting that feedback timing may have little effect on L2 vocabulary acquisition in flashcard learning conditions.

Criteria 16 and 17 in TFA ask whether the activity involves instantiation and imaging, respectively. In instantiation, target words are used in a meaningful, visual situation. When the meanings of target words are illustrated visually, the activity involves imaging. Although flashcards can include example sentences or pictures, they usually do not promote instantiation or imaging. As a result, flashcard learning meets neither Criterion 16 nor 17. Due to their multimedia capabilities, computer-based flashcards may have some potential to facilitate instantiation and imaging (Nakata, 2011).

The last criterion in TFA asks whether the activity is designed to avoid interference. Studies have shown that learning semantically related words such as synonyms or antonyms together has a negative effect on L2 vocabulary learning because it tends to cause interference between words (for a review, see Nation, 2000). Unless learners study semantically related words together, interference can be avoided using flashcards. Flashcard learning, therefore, sometimes meets this criterion. However, it should be noted that some vocabulary learning computer programs offer sets of readymade flashcards that are semantically related (e.g., words related to food, animals, or colors). Because some learners, teachers, and materials developers tend to believe that semantically related words should be studied together (Folse, 2004), it may be useful to raise awareness that interference inhibits L2 vocabulary learning.

Let us now look at the other four guidelines proposed by Nation (2013a). Flashcard learning typically does not require a large amount of work from teachers because flashcards are usually prepared by students. Various kinds of readymade flashcards, both paper- and computer-based, are also available. Regarding the four strands, flashcard learning is typically used for language-focused learning and does not provide opportunities for meaning-focused input, meaning-focused output, and fluency development. It may be useful, therefore, for teachers and curriculum developers to ensure that L2 words studied using flashcards are also met in the other three strands. Flashcard learning satisfies Nation's guideline of efficiency because a large number of words can be learned using flashcards in a short amount of time (e.g., Fitzpatrick, Al-Qarni, & Meara, 2008; Nation, 2013b). The activity also meets Nation's fifth guideline because it can be used many times. If the same words are practiced more than once, it may be desirable to practice receptive retrieval first and then productive retrieval (e.g.,

Nation, 2013b).

Cloze Exercises

The next activity to be analyzed is the cloze exercise, where learners are asked to complete sentences by filling in blanks. Cloze exercises can be done in two ways: in one, students have to fill in blanks from memory (recall), in the other, students fill in blanks based on a list of possible choices (recognition). In our example, we are using the former format. An example is given below:

Figure 8.1 *Example of a cloze exercise*

- | |
|---|
| <ol style="list-style-type: none">1. This building is made of (). セメント2. The () of living is expensive in big cities. 費用3. I buy my vegetables at the () every week on Sunday morning. 市場 |
|---|

Note. The correct answer is *cement* for 1, *cost* for 2, and *market* for 3. Adapted from Webb (2012, p. 130).

In the above example, L1 (Japanese) translations of the missing words are also given (セメント, 費用, and 市場) in order to prevent learners from supplying words that are acceptable in the sentence but are different from target words that we want them to practice. For instance, some learners may produce *shop* instead of the target word *market* for the third question. It may also be useful to provide the first letter (e.g., m____) and / or the number of letters of in the word (e.g., six letters) instead of L1 translations.

Now let us consider the effectiveness of this activity. Table 8.2 (middle) indicates the efficacy of cloze exercises using TFA.

Table 8.2 Analysis of cloze exercises and crossword puzzles using Technique

Feature Analysis

		Cloze exercises		Crossword puzzles	
	Criteria	Score	Comment	Score	Comment
	Motivation				
1	Vocabulary learning goal	1		1	
2	Motivate learning	1	Yes, because it is challenging.	1	Yes, because some learners find puzzles interesting.
3	Words selected by learners	0	No, but it could involve self-selected words (see text).	0	
	Noticing				
4	Attention on target words	1		1	
5	Awareness of new vocabulary learning	1		1	
6	Negotiation	0		0	No, but it could involve negotiation (see text).
	Retrieval				
7	Retrieval	1	Yes, because learners have to retrieve L2 word forms.	1	Yes, because learners have to retrieve L2 word forms.
8	Productive retrieval	1	Yes, because learners have to retrieve L2 word forms.	1	Yes, because learners have to retrieve L2 word forms.
9	Recall	1	Yes, because learners have to recall L2 word forms.	1	Yes, because learners have to recall L2 word forms.
10	Multiple retrievals	0	No, but it could involve multiple retrievals (see text).	0	No, but it could involve multiple retrievals (see text).
11	Spacing between retrievals	0	No, but it could involve spacing (see text).	0	No, but it could involve spacing (see text).
	Generation				
12	Generation	1	Yes, because learners meet L2 words in	0	No, but it could involve generation (see text).

			novel contexts.		
13	Productive generation	0	No, because learners meet, not use, L2 words in novel contexts.	0	
14	High degree of generation	0		0	
	Retention				
15	Successful form-meaning linking	0	No, but successful form-meaning linking could be ensured (see text).	0	No, but successful form-meaning linking could be ensured (see text).
16	Instantiation	0		0	
17	Imaging	0		0	
18	Avoidance of interference	1	Yes, unless semantically related words are practiced together.	0	No, because semantically related words (names of vegetable) are practiced together in our example. But it could be improved (see text).
	Total score	9		7	

Note. Adapted from Nation and Webb (2011, pp. 318-319).

Table 8.2 shows that cloze exercises have a TFA score of 9 out of 18. This is the second highest among 12 exercises analyzed by Nation and Webb (2011), suggesting that cloze exercises are a relatively effective vocabulary learning technique. At the same time, Table 8.2 also shows that the activity could be improved, especially in the retrieval and retention categories. One way to modify the activity would be to give multiple questions for a given target word (Folse, 2006). For instance, the target word *cement* could be practiced in three cloze questions such as the following:

Figure 8.2 Example of cloze questions activity with target word ‘cement’

1. This building is made of (). セメント
2. The () of living is expensive in big cities. 費用
3. We () the path. 舗装した
4. I buy my vegetables at the () every week on Sunday morning. 市場
5. We () our friendship with a drink. 強固にした
6.

Note. The correct answer is *cement* for 1, *cost* for 2, *cemented* for 3 and 5, and *market* for 4. セメント is the Japanese translation for the noun *cement*. 舗装した is the Japanese translation for the past tense of the verb *cement* when used literally. 強固にした is the Japanese translation for the past tense of the verb *cement* when used figuratively. Adapted from Nation (2013, p. 110) and Webb (2012, p. 130).

Because the activity shown above involves three retrievals of the target word *cement*, it now meets Criterion 10 (multiple retrievals). When giving multiple sentences for a given target word, it is useful to ensure that sentences for a given target word are separated by those for other words. For instance, in the above example, *cement* is practiced in Questions 1, 3, and 5, not in 1, 2, and 3. By doing so, the activity now introduces spacing between retrievals and satisfies Criterion 11. Note that the sentences are designed to promote generation. In Question 1, *cement* is a noun and used in its most frequent, basic meaning. In Question 3, the word is used as a verb and takes a concrete noun (*path*) as the object. In Question 5, the word is used as a verb, takes an abstract noun (*friendship*) as the object, and is used figuratively. Although

these features may be useful, they do not contribute to the TFA score because the activity is already given a point for generation (Criterion 12).

Another way to increase the learning potential of cloze exercises would be to give correct answers. By doing so, we can ensure that learners have opportunities to make the correct form-meaning connection, and the activity will receive a point for Criterion 15 (successful form-meaning linking). If cloze exercises are given in a computer program, feedback can be provided relatively easily (for an example, see <http://www.lex tutor.ca/concordancers/multi/>).

Motivational aspects of cloze exercises could be improved by allowing learners to practice self-selected target words. Multi-Concordance (<http://www.lex tutor.ca/concordancers/multi/>) created by Tom Cobb enables learners to do this. If learners type L2 words, the software automatically generates cloze exercises for each word based on sentences extracted from a corpus. Below is an example of cloze exercises created by the software for the word *cement*:

Figure 8.3 Example of cloze exercises based on Brown Corpus created by

Tom Cobb's Multi-Concordance (<http://www.lex tutor.ca/conc/multi/>).

< INSERT FIGURE 8.3 HERE (please take screenshot of dialog box in

<http://lex tutor.ca/cgi-bin/conc/multi.pl?inputs=cement&this corp=brown strip.txt&limit=12&wid=large&format=want quiz&dico=Eng Eng&language=&cumul score=>) >

If all of the above changes are made, the activity will receive a total TFA score of 13, which is higher than any other activities analyzed by previous research so far (Nation & Webb, 2011; Webb, 2012).

Let us now look at the other four guidelines proposed by Nation (2013a). Cloze exercises require a reasonable amount of work from teachers because the exercises need to be created. It is true that software such as Multi-Concordance allows teachers to create cloze exercises automatically. However, exercises generated by software may need to be edited by teachers because they might contain a large number of low-frequency words and may not necessarily be suitable for learners. Regarding the four strands, cloze exercises are mainly concerned with language-focused learning and do not provide opportunities for the other three strands. Cloze exercises satisfy Nation's guideline of efficiency because the task can be completed relatively quickly. At the same time, because the task requires more time than learning from flashcards, it may be less efficient than flashcard learning. The activity also meets Nation's fifth guideline because it can be used many times. However, when the same target words are practiced in cloze exercises, it may be desirable not to use the same contexts repeatedly because it may not be very motivating.

Crossword Puzzles

The next activity to be analyzed is the crossword puzzle, where learners produce L2 words based on the meaning given. An example is given below:

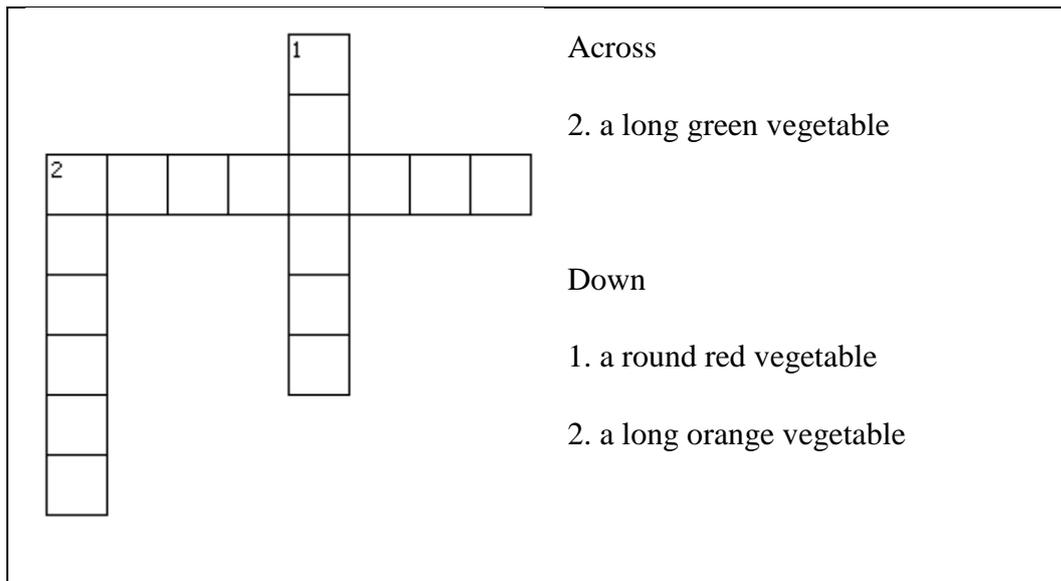


Figure 8.4 Example of a crossword puzzle. The correct answer is *cucumber* for 2 Across and *tomato* and *carrot* for 1 and 2 Down, respectively. Adapted from Webb (2012, p. 129).

Created with *Puzzlemaker*

(<http://puzzlemaker.discoveryeducation.com/CrissCrossSetupForm.asp>).

First, we will consider the effectiveness of this activity. Table 8.2 (right) indicates the efficacy of crossword puzzles using TFA. Table 8.2 shows that crossword puzzles have a TFA score of 7 out of 18 and may be less effective than flashcards (12) or cloze (9). Note that the activity receives no point for the generation and retention categories. Let us see how this activity could be improved to increase the learning potential. First, by having students solve crosswords in a pair or group instead of individually, the activity may involve negotiation of word meanings and satisfy Criterion 6 of TFA.

Another way to modify the activity would be not to practice semantically related words (e.g., names of vegetable) together. For instance, instead of *carrot*, *tomato*, and *cucumber*, target words such as *carrot*, *cost*, and *market* could be practiced. This would avoid interference, and

used as a verb in 2 Across, it is used as a noun in 2 Down. This facilitates generation (Criterion 12). Using inflected forms (e.g., *costing*, *carrots*, and *markets* instead of *cost*, *carrot*, and *market*) as answers may also help promote generation. If all of the above changes are made, the activity will receive a total TFA score of 13, the same score as the revised cloze exercises.

One potential problem with crossword puzzles, though, is that learners may be able to solve some questions without carefully considering the cues. For instance, suppose that learners solved 2 Across (*cost*) and 3 Across (*carrot*) before 2 Down in the above example. By solving these questions first, learners may be able to infer that the correct answer for 2 Down is a four-letter word that starts with *c* and ends with *t*. As a result, learners may be able to come up with the correct answer for 2 Down (*cost*) without carefully examining the definition or cloze sentence. This is not very effective because it may deprive learners of opportunities to strengthen form-meaning connection or study how L2 words are used in context.

Let us now look at the other four guidelines proposed by Nation (2013a). Because creating crossword puzzles may require a substantial amount of work from teachers, the activity does not meet Nation's (2013a) second guideline. However, a web site such as *Puzzlemaker* (<http://puzzlemaker.discoveryeducation.com/CrissCrossSetupForm.asp>) can help teachers to create crossword puzzles in a relatively short time. Regarding the four strands, crossword puzzles are mainly concerned with language-focused learning and do not provide opportunities for the other three strands. Crossword puzzles may be less efficient than flashcard learning and cloze exercises because they take more time. The activity meets Nation's fifth guideline because it can be used many times.

Discussion

In this chapter we looked at the relative efficacy of three common vocabulary learning activities using guidelines proposed by Nation (2013a). The analysis found that Nation's guidelines may be useful for highlighting possible strengths and weaknesses of vocabulary learning activities. For instance, the analysis revealed that although flashcard learning may be very effective in terms of retrieval, its ability to promote generation may be limited. In contrast, although crossword puzzles may be motivating, they may not be very effective in terms of generation and retention, at least unless some modifications are made. Flashcard learning may have additional value because it requires less work from teachers and time on task compared with cloze exercises and crossword puzzles. Examining activities using these guidelines may be useful because it may allow teachers and materials developers to determine which activities might be used based on learning goals or learners' needs. Teachers and materials developers could also combine different activities so that they can complement each other's weaknesses. For instance, by practicing same words in flashcards first and then revised cloze exercises, learners may be able to benefit from the positive effects of both retrieval and generation. The chapter also showed how TFA may be used to evaluate and indicate where modification of vocabulary learning activities may be useful to increase learning potential. For example, by making modifications to cloze exercises and crossword puzzles, both activities might be made to be more effective by including generation and retrieval within the activity.

Present and previous analyses using TFA (Nation & Webb, 2011; Webb, 2012) also reveals that few vocabulary learning activities involve instantiation or imaging. For instance, none of the activities analyzed in the present and previous analyses promotes instantiation. Similarly,

only one activity (keyword technique; Nation & Webb, 2011) analyzed so far facilitates imaging. This suggests that it may be useful for teachers and materials developers to consider instantiation and imaging when designing vocabulary learning activities. Including an instantiation or imaging component within an activity may be particularly useful for developers of CALL materials, where instantiation and imaging can be promoted relatively easily because of multimedia capabilities (WUFUN provides one example of how this might be done; Ma & Kelly, 2006).

As this chapter has shown, Nation's (2013a) framework may be very useful for evaluating and indicating where modification is necessary within vocabulary learning activities. Further analysis of more activities using this framework would be useful. Teachers and materials developers may then be able to use this information to improve the efficacy of vocabulary learning programs.

References

- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*, 35–56.
doi:10.1111/j.1467-9922.2007.00398.x
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*, 273–281. doi:10.1037/1076-898X.13.4.273
- 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.
doi:10.1080/09541440701326097
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition, 34*, 268–276. doi:10.3758/BF03193405
- 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
- Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal, 36*, 239–248. doi:10.1080/09571730802390759
- Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*, 273–293. doi:10.2307/40264523
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.),

Cognition and second language instruction (pp. 258–286). Cambridge, UK:
Cambridge University Press.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*, 357–377.

doi:10.1093/applin/19.3.357

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (in press). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0636-z

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558. doi:10.1080/09541440601056620

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250–1257. doi:10.1037/a0023436

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968. doi:10.1126/science.1152408

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *The Modern Language Journal*, *73*, 440–464.

doi:10.1111/j.1540-4781.1989.tb05325.x

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97. doi:10.3102/00346543058001079

Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation.

Computer Assisted Language Learning, *19*, 15–45. doi:10.1080/09588220600803998

Metcalf, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in

children's and adults' vocabulary learning. *Memory & Cognition*, 37, 1077–1087.

doi:10.3758/MC.37.8.1077

Mondria, J.-A., & Mondria-de Vries, S. (1994). Efficiently memorizing words with the help of word cards and “hand computer”: Theory and applications. *System*, 22, 47–57.

doi:10.1016/0346-251X(94)90039-6

Mondria, J.-A., & Wiersma, B. (2004). Receptive, productive and receptive + productive L2 vocabulary learning: What difference does it make? In B. Laufer (Ed.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 79–100). Amsterdam, Netherlands: John Benjamins.

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

Nakata, T. (in press). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*.

doi:10.1177/1362168814541721

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning.

ReCALL, 20, 3–20. doi:10.1017/S0958344008000219

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

Nakata, T. (2013a). Web-based lexical resources. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 6166–6177). Oxford, UK: Wiley-Blackwell.

Nakata, T. (2013b). *Optimising second language vocabulary learning from flashcards*

- (Unpublished doctoral dissertation). Victoria University of Wellington, New Zealand.
- Nation, I. S. P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, 9, 6–10.
- Nation, I. S. P. (2013a). What are the ten most effective vocabulary activities? Presented at the JALT College and University Educators Special Interest Group Event, Sapporo, Japan.
- Nation, I. S. P. (2013b). *Learning vocabulary in another language* (2nd edition.). Cambridge, UK: Cambridge University Press.
- Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle Cengage Learning.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1051–1057. doi:10.1037/0278-7393.29.6.1051
- Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, 35, 1917–1927. doi:10.3758/BF03192925
- Rohrer, D., Taylor, K. M., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, 19, 361–374. doi:10.1002/acp.1083
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, 29, 449–484. doi:10.1017/S0272263107070271

Van Bussel, F. J. J. (1994). Design rules for computer-aided learning of vocabulary items in a second language. *Computers in Human Behavior*, 10, 63–76.

doi:10.1016/0747-5632(94)90029-9

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

Webb, S. A. (2012). Learning vocabulary in activities. In H. P. Widodo & A. Cirocki (Eds.), *Innovation and creativity in ELT methodology* (pp. 121–133). New York: NY: Nova.

¹ It should be noted that most activities can be done in several different ways. We have selected commonly used formats for these three activities. However, it would be fair to argue that we are analyzing a particular way of using each of the three activities, and other manifestations of the activities may have slightly different strengths and weaknesses.