20

Learning Words With Flash Cards and Word Cards

Tatsuya Nakata

Introduction

Flash cards (sometimes referred to as word cards) are a set of cards where the second language (L2) word is written on one side and its meaning, usually in the form of a first language (L1) translation, L2 synonym, or definition, is written on the other. Flash card learning is a type of paired-associate learning, where target items are encountered in a decontextualized format, and learners are asked to connect the L2 word form and its meaning. Possibly due to its unfortunate associations with behaviorism, flash card learning is not necessarily popular among researchers and teachers. Existing studies, nonetheless, show that flash card learning is a common learning strategy. Wissman, Rawson, and Pyc (2012), for instance, surveyed 374 American undergraduate students and report that 67.6% of them study with flash cards. Similarly, Schmitt (1997) surveyed 600 Japanese learners of English and found that 51% of junior high school students and 29% of senior high school students use flash cards for learning English vocabulary. A number of flash card programs are also available as computer software, smartphone applications, or web-based applications (for a review, see Nakata, 2011).

Research also shows that flash card learning is not only common but also effective and efficient. Studies demonstrate that memory for paired-associates may persist over years if reviewed regularly (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987). Although contextual vocabulary learning is often considered superior to paired-associate learning, research shows that paired-associate learning may be as effective as or more effective than vocabulary learning from context (Laufer & Shmueli, 1997; Prince, 1996; Webb, 2007). A recent review also found that despite the belief that memory of paired-associate learning decays quickly, it may be as effective as the keyword mnemonic in the long term (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Flash card learning is also efficient because it allows learners to acquire a large number of lexical items in a relatively short time (Fitzpatrick, Al-Qarni, & Meara, 2008; Nation, 1980).

Studies also suggest that flash card learning improves L2 learners' ability to use the words in communication and is a useful learning activity. Elgort (2011), for instance, showed that flash card learning contributes to implicit as well as explicit vocabulary knowledge, thus resulting in not only the *learning* but also the *acquisition* of L2 vocabulary. Webb (2009a) also found that paired-associate learning improves learners' ability to use L2 words in both reading and writing. Although paired-associate learning is only an initial step towards mastering new vocabulary, most researchers agree that it should have a place in L2 vocabulary instruction. Hulstijn (2001) and Laufer (2005), for instance, point out that incidental vocabulary learning alone is insufficient and should be supplemented with word-focused activities such as paired-associate learning. Nation (2008) considers flash card learning to be one of the most effective vocabulary learning strategies, along with guessing from context, word part learning, and dictionary use. Nation and Webb (2011) analyzed 12 vocabulary learning activities using a framework called Technique Feature Analysis and concluded that flash card learning is the most effective. Of course, flash card learning is not without limitations and should only be a part of a well-balanced curriculum (Nation, 2013). At the same time, given the effectiveness, efficiency, and usefulness of flash card learning, this chapter reviews empirical studies on flash card learning and discusses principles for effective flash card learning. Note that where necessary, this chapter discusses not only studies on flash card learning but also studies on paired-associate learning in general. This is because we can expect the findings of paired-associate learning studies to be applicable to flash card learning.

Critical Issues and Topics

Effects of Retrieval on Flash Card Learning

Perhaps one of the most important principles for effective flash card learning is to use retrieval. Retrieval is defined as remembering information about previously encountered L2 words. For example, suppose English speakers are attempting to learn the Japanese word *inu* (dog). When learners translate *inu* into English, this requires them to recall its meaning (dog). Hence, this activity involves retrieval. When the Japanese word *inu* is presented together with its meaning (dog), however, learners do not need to recall any information about *inu*. This activity, therefore, does not include a retrieval component. Studies have shown that retrieval significantly increases L2 vocabulary learning (e.g., Barcroft, 2007; Karpicke & Roediger, 2008).

Karpicke and Roediger (2008), for instance, examined the effects of retrieval on L2 vocabulary learning. In their study, 40 American undergraduate students studied Swahili-English word pairs (e.g., *malkia – queen*) using flash card software. The treatment consisted of a series of study (presentation) and test (retrieval) episodes. In a study episode, participants were presented with a Swahili word together with its L1 (English) translation (e.g., *malkia – queen*). In a test episode, participants were presented with a Swahili word and asked to type in the corresponding L1 (English) translation (e.g., *malkia* = _____?). No correct response was provided as feedback after each test episode. Learning was measured by a one-week delayed posttest, where participants were asked to translate Swahili words into English. The posttest results showed that while increasing the number of study episodes had little effect on retention (33 → 36%), increasing the number of test episodes increased retention significantly (33 → 81%). Their findings suggest that it is the test episodes (retrievals), not study episodes, that determine retention.

One important question is whether the benefits of repeated retrieval are retained in the long term: Karpicke and Roediger (2008) did not administer a posttest after a delay greater than one week. In an L1 vocabulary study, Rohrer, Taylor, Pashler, Wixted, and Cepeda

(2005) showed that although practicing retrieval repeatedly facilitates learning in the short term, it is not necessarily beneficial in the long term. In Rohrer et al.'s study, 88 American undergraduate students studied low-frequency English words paired with higher frequency synonyms (e.g., *acrogen – fern*). The participants were divided into two groups: low and high. The low group practiced retrieval of each word pair (e.g., *acrogen* = _____?) five times, and the high group practiced retrieval ten times. While the high group fared significantly better than the low group on a one-week delayed posttest (low: 38%; high: 64%), no statistically significant difference existed between the two groups four weeks after the treatment (low: 18%; high: 22%). Rohrer et al.'s study indicates that the positive effects of repeated retrieval may disappear over time.

In order to test whether Rohrer et al.'s (2005) findings are also applicable to L2 vocabulary learning, Nakata (2017) attempted to assess the long-term effects of repeated retrieval. In his study, 98 Japanese college students studied 16 English-Japanese word pairs using flash card software. The participants were randomly assigned to one of the four retrieval frequency levels: one, three, five, and seven. Nakata found that five and seven retrievals were significantly more effective than one and three retrievals not only immediately and one week later, but also four weeks after the treatment. The results indicate that contrary to the findings of Rohrer et al. (2005), the advantage of repeated retrieval may be retained at least four weeks after learning. Repeated retrieval, however, inevitably increased the study time. For instance, while practicing retrieval of 16 target items once took only 6.48 minutes, practicing the same number of items seven times required 22.82 minutes. When the number of words learned per minute was calculated, Nakata found that single retrieval resulted in larger gains than repeated retrieval. The findings suggest that (1) if learners have enough time, increasing retrieval frequency is desirable, and (2) if learners do not have enough time, it may be efficient to practice retrieval only once.

Research also shows that the direction of retrieval (i.e., receptive vs. productive) affects learning. Productive retrieval is defined as remembering the L2 word corresponding to the meaning, while receptive retrieval is defined as remembering the meaning corresponding to the L2 word. For instance, when English speakers are asked to translate *dog* into Japanese, this involves productive retrieval. In contrast, when English speakers are asked to translate the Japanese word *inu* (dog) into English, this involves receptive retrieval. (Some researchers prefer to use the term *meaning recall* instead of receptive retrieval and *form recall* instead of productive retrieval.) Research shows that receptive retrieval is effective if learning is measured by a receptive test, whereas productive retrieval is effective if learning is measured by a productive test (e.g., Steinel, Hulstijn, & Steinel, 2007; Webb, 2009a, 2009b). The results are consistent with the transfer-appropriate processing framework (Morris, Bransford, & Franks, 1977), which states that learning is most optimal when the learning condition matches the testing condition. The findings suggest that if learning both productive and receptive lexical knowledge is the goal, learners should practice retrieval both productively and receptively. Research also shows that while receptive retrieval is not very effective for learning productive knowledge, productive retrieval results in a relatively large increase in receptive knowledge (e.g., Schneider, Healy, & Bourne, 2002; Steinel et al., 2007; Webb, 2009a). The findings suggest that if learners have to choose either receptive or productive retrieval for some reason (e.g., they do not have enough time to practice both directions), it may be useful to practice productive retrieval rather than receptive retrieval.

In addition to the dichotomy of receptive vs. productive, retrieval can also be categorized according to the dichotomy of recognition vs. recall. In recall, learners are required to generate a response, whereas in recognition, learners are required to select a correct response from options provided. For instance, when learners are asked to translate *dog* into Japanese, the treatment involves recall. In contrast, when learners are asked to choose the Japanese equivalent of *dog* from multiple-choice options, this involves recognition. The dichotomies of receptive vs. productive and recognition vs. recall result in the following four kinds of retrieval formats: receptive recognition, receptive recall, productive recognition, and productive recall (e.g., Laufer & Goldstein, 2004). One may wonder which of these four formats is the most effective for learning. Nakata (2016) found that the answer probably depends on how vocabulary knowledge is measured. His study showed that if learning was measured by a productive recall test, learning using productive recall was the most effective, which is consistent with the transfer-appropriate processing framework. However, learning using recognition was more desirable than learning using recall when measured by one of the following three tests: receptive recall, productive recognition, and receptive recognition. Recognition formats also required less study time than recall formats. Pedagogically, the findings suggest that (1) for learning the productive knowledge of orthography, productive recall should be used, and (2) when the productive knowledge of orthography is not the goal, recognition formats should be used.

Types of Information

One important decision that learners have to make in flash card learning is what kind of information to add to the flash cards. Because research shows that the use of L1 translations increases vocabulary learning (e.g., Laufer & Shmueli, 1997), it may be useful to provide the meaning in the form of L1 translations rather than L2 definitions or synonyms. Research also suggests that providing the core meaning of polysemous words facilitates the acquisition of multiple meanings of words (Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). The core meaning refers to the most central, underlying meaning of a word. For instance, the word *draw* can be used in many senses, as in *draw a picture*, *draw a curtain*, *draw criticism*, *draw attention*, *draw a bath*, or *open a drawer*. Most dictionaries give different translations or definitions for each of these senses. For instance, in a typical English-Japanese dictionary, *draw* in *draw a picture* is translated into 描く, while *draw* in *draw a curtain* is translated into 引く. However, most senses of *draw* are related to its core meaning, which is to move or pull something. For instance, *draw a picture* involves making a line by moving a pen. *Draw attention* or *draw criticism* involves bringing (or moving) attention or criticism to somebody. A *drawer* is something you open by pulling. Providing the core meaning may be useful because it helps learners to understand how different meanings of a word are connected to each other.

Although flash card learning is often considered as a form of decontextualized learning, it is possible to add short contextual sentences to flash cards. Empirical studies, however, suggest that context has no or little effect on vocabulary learning (Laufer & Shmueli, 1997; Prince, 1996; Webb, 2007). It may also be useful to add some information that helps connect the form and meaning of L2 words such as the keyword (e.g., Ellis & Beaton, 1993; Hulstijn, 1997; Rodriguez & Sadoski, 2000) or word parts (Wei, 2015). The keyword refers to a word in the L1 that sounds like an L2 target word. For instance, when learning the French word *clef* (key), it may be useful to think of an English word that sounds like *clef* such as *cliff* and create a mental image that connects the two (e.g., a key is on top of a cliff; Dunlosky et al., 2013). Even if learners forget the meaning of *clef*, the keyword *cliff* and the mental image may help learners to retrieve its meaning. Adding information about word parts to flash cards is also useful because the knowledge of affixes and roots may help strengthen form-meaning connection. For instance, the English word *prospect* can be broken down into *pro* and *spect*, where *pro* means *forward* (as in *proceed* or *progress*) and *spect* means *to look* (as in *spectacle* or *spectator*). Knowing the meanings of *pro* and *spect* may help learners to remember the meaning of *prospect* (i.e., *prospect* is a looking forward, so it means *anticipation* or *expectation*; Wei, 2015). In computer-based flash cards, multimedia information such as images, videos, or audios can also be added to flash cards. Images and videos are useful because visual information helps illustrate the meaning of words (especially in the case of concrete nouns or verbs describing actions) and may facilitate learning (e.g., Carpenter & Olson, 2012). Adding audio recordings of L2 words may also be useful for acquiring their spoken form.

One caveat to be considered is that trying to teach or learn multiple aspects of unfamiliar words at once may lead to information overload and possibly have negative effects on learning. Barcroft (2002), for instance, found that when study time is limited (e.g., 12 seconds per target word), although paying attention to the word form (structural elaboration) facilitates the acquisition of knowledge of the word form, it inhibits the acquisition of knowledge of the meaning. Similarly, Barcroft showed that although paying attention to the meaning (semantic elaboration) facilitates the acquisition of meaning, it inhibits the acquisition of knowledge of the word form. These findings are known as the Type of Processing – Resource Allocation (TOPRA) model (Barcroft, 2002, 2015). Because the knowledge of form-meaning connection is considered the most important aspect, it is perhaps best to focus on linking meaning with form in the first encounter with unfamiliar words and try to acquire other aspects (e.g., spelling, collocations, constraints on use) only in later stages.

Effects of Practice Schedule

Even though flash card learning is a very effective vocabulary learning method, memory decays over time and needs to be reinforced. This raises the question of what kind of review schedule should be used to maximize long-term retention. The review schedule can be divided into massed learning and spaced learning. In massed learning, the same materials are studied multiple times sequentially (e.g., studying the Japanese word *inu* repeatedly in a row). In spaced learning, the same materials are studied multiple times after certain intervals (e.g., studying the Japanese word *inu* every week). Research shows that spaced learning increases L2 vocabulary learning more than massed learning (e.g., Nakata, 2015a). This finding is known as the spacing effect. Studies have also shown that the magnitude of the spacing effect tends to be large. Nakata (2015a), for instance, found that a spaced schedule increased learning by more than 100% compared to a massed schedule on a one-week delayed posttest (massed = 14%; spaced = 29%).

Since spaced learning is more effective than massed learning, one might wonder how retrieval opportunities should be distributed in order to maximize retention. Studies have demonstrated that longer lags between retrievals often facilitate long-term retention more than shorter lags (e.g., Karpicke & Bauernschmidt, 2011; Nakata & Webb, 2016a). For instance, practicing retrieval of a given item every week (lag of one week) facilitates retention more than practicing retrieval every day (lag of one day). This finding is referred to as the lag effect. While spaced learning is found to be more effective than massed learning irrespective of the retention interval, the lag effect is sometimes not observed depending on the retention interval. (The retention interval refers to an interval between the last encounter with a given item and the posttest. For instance, if the posttest is given ten days after the treatment, the retention interval is ten days.) Specifically, while long lags between retrievals are often effective when learning is measured at a long retention interval, short lags are sometimes effective when learning is measured at a short retention interval. This finding is referred to as the spacing-by-retention interval interaction (e.g., Nakata, 2015a; Nakata & Webb, 2016a). Research suggests that L2 vocabulary acquisition is influenced by the spacing effect, the lag effect, and the spacing-by-retention interval interaction (e.g., Bahrick et al., 1993; Nakata, 2015a; Nakata & Webb, 2016a).

In Nakata (2015a), for instance, 128 Japanese university students studied 20 English-Japanese word pairs using flash card software. The participants were randomly assigned to one of the following four spacings: massed-, short-, medium-, and long-spaced. The participants in the massed group practiced retrieval of a given English word three times in a row. The short-, medium-, and long-spaced groups used average spacing of approximately one, two, and six minutes, respectively (that is, participants in the long-spaced group practiced retrieval of a given target word approximately every six minutes). Nakata found that the massed group was the least effective among the four groups both immediately and one week after the treatment, supporting the spacing effect. Moreover, while the short-spaced group fared significantly better than the massed group immediately after learning, the advantage disappeared at a one-week retention interval. The medium- and long-spaced groups, in contrast, were significantly more effective than the massed group not only on the immediate but also on the one-week delayed posttest. The findings are consistent with the spacing-by-retention interval interaction, which states that the benefits of short lags disappear over time. Nakata and Webb (2016a, Experiment 2) also observed the positive effects of long lags for L2 vocabulary learning. They compared the effects of short and long spacing. Encounters of a given target item were separated by approximately 30 seconds and three minutes in the short- and long-spaced conditions, respectively. While the long-spaced group failed to outperform the short-spaced group immediately after the treatment (short = 58%; long = 62%), the long-spaced group fared significantly better than the short-spaced group at a one-week retention interval (short = 8%; long = 20%; Figure 20.1). These findings support the lag effect and spacing-by-retention interval interaction.

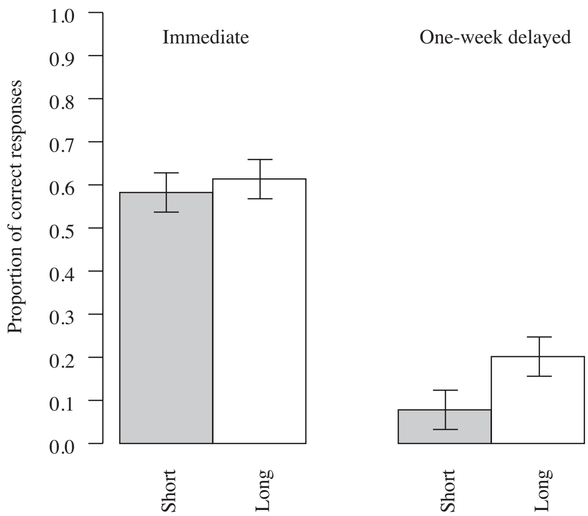

Figure 20.1 Proportion of correct responses on the immediate and one-week delayed posttests in Nakata and Webb (2016a, Experiment 2)

Note: The error bars indicate standard errors.

Source: Adapted from Nakata and Webb, 2016a, Experiment 2

One important finding related to spacing is that although long lags tend to induce more retrieval failures during the learning phase compared with no or short lags, long lags often lead to superior retention in the long term. For instance, in Nakata (2015a), although the spaced group produced significantly fewer correct responses during the learning phase than the massed group (massed = 89%; spaced = 21%), the results were reversed on the posttest conducted immediately after learning (massed = 32%; spaced = 52%). On the one-week delayed posttest, the spaced group's average score was more than twice as high as that of the massed group (massed = 14%; spaced = 29%; see Figure 20.2). Similarly, in Nakata and Webb (2016a, Experiment 2), although the short-spaced group produced more correct responses during the learning phase (short: 76%; long: 44%), the long-spaced group significantly outperformed the short-spaced group one week after the treatment (short: 8%; long: 20%). These findings are consistent with the desirable difficulty framework (Bjork, 1999), according to which a treatment that increases learning difficulty can be effective in the long term. Pedagogically, the findings suggest that teachers and learners should not be concerned too much about learning phase performance. Methodologically, the findings suggest that it may be inappropriate to equate learning phase performance with long-term retention (however, see Tinkham, 1993, 1997; Waring, 1997).

Another important finding related to spacing is that learners are often unaware that spacing increases learning. In Nakata and Suzuki (in press), for instance, 133 Japanese college students studied 48 English-Japanese word pairs using flash card software. Half of the participants were assigned to the short-spaced group, while the other half were assigned to the long-spaced group. Encounters of a given target item were separated by five and 40 trials on average in the short- and long-spaced groups, respectively. Learning was measured by a posttest conducted immediately after and one week after the treatment. Upon completion of the immediate posttest, learners in both groups were asked to predict how many target

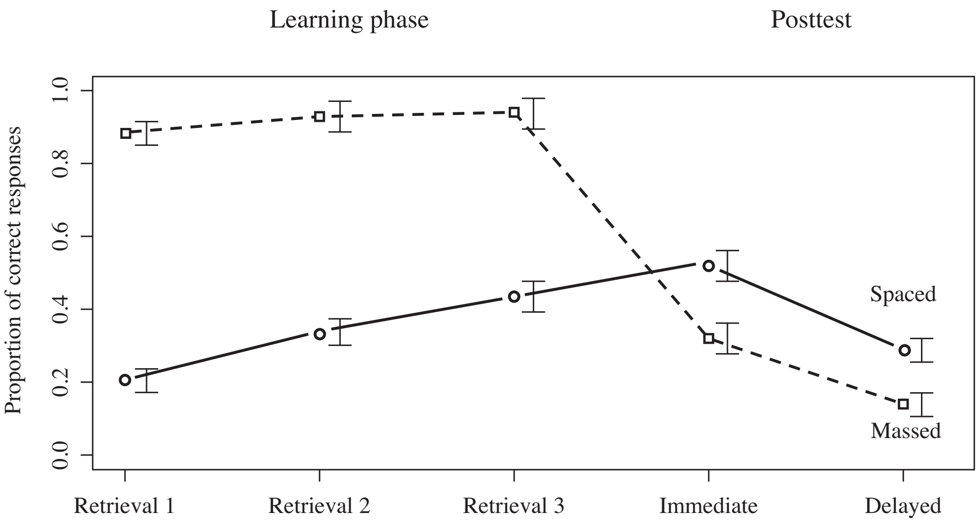

Figure 20.2 Learning phase and posttest performance in Nakata (2015a). The three data points on the left show the proportion of correct responses during the learning phase, and the two data points on the right show the proportion of correct responses on the posttests.

Note: The error bars indicate standard errors.

Source: Adapted from Nakata, 2015a

items out of 48 they might be able to recall in one week (judgments of learning). Although the long-spaced group significantly outperformed the short-spaced group on the one-week delayed posttest (short: 17%; long: 27%), the short-spaced group predicted that they would recall as many target words as the long-spaced group (short: 29%; long: 28%), overestimating retention (Figure 20.3). The findings are consistent with metacognition literature showing that learners are often unaware that spacing increases learning (e.g., Kornell, 2009; Wissman et al., 2012). The results support raising awareness of the value of spacing for vocabulary learning.

Spacing schedules can also be divided into expanding and equal spacing. Expanding spacing refers to a practice schedule where the lags between retrieval opportunities of a given word are increased in an expanding manner (e.g., two weeks, four weeks, six weeks). Equal spacing refers to a schedule where the lags between retrievals of a given word are fixed (e.g., four weeks, four weeks, four weeks). While many researchers claim that expanding spacing is the most optimal review schedule (e.g., Barcroft, 2015; Hulstijn, 2001; Nation & Webb, 2011), empirical evidence supporting the value of expanding spacing for L2 vocabulary learning is scant (Kang, Lindsey, Mozer, & Pashler, 2014; Karpicke & Bauernschmidt, 2011; Nakata, 2015a; Pyc & Rawson, 2007). Karpicke and Bauernschmidt (2011), for instance, did not find any significant difference between the effects of equal and expanding spaced conditions. In Nakata (2015a), 96 Japanese students studied 20 English-Japanese word pairs using flash card software. Half of the target items were studied under equal spaced conditions, while the other half were studied under expanding spaced conditions. Posttest results

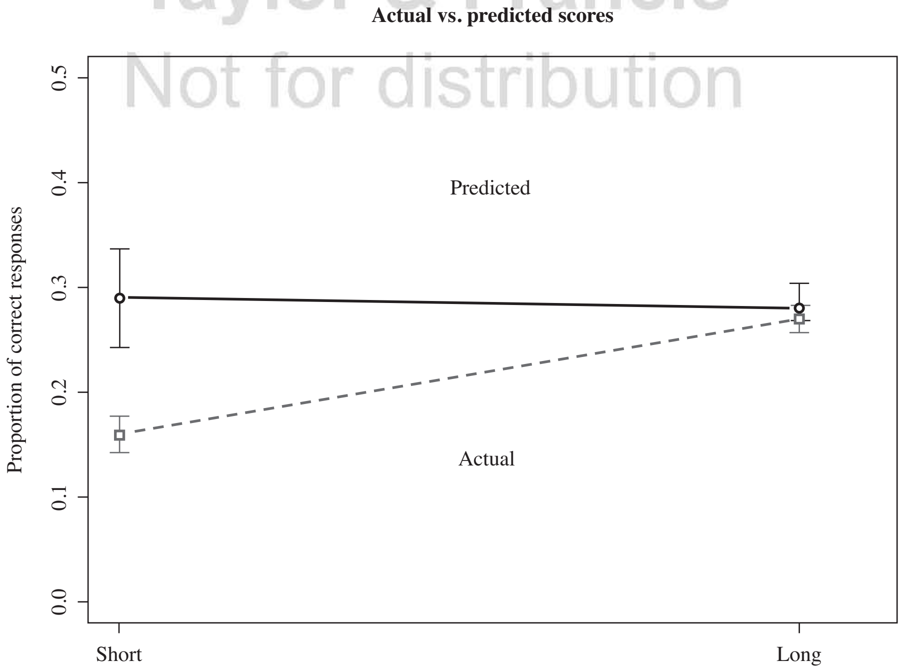

Figure 20.3 Actual vs. predicted scores on the one-week delayed posttest in Nakata and Suzuki (in press)

Note: The error bars indicate standard errors.

Source: Adapted from Nakata and Suzuki, in press showed only limited, if any, superiority of expanding spacing over equal spacing. Overall, existing studies indicate that although introducing long lags between retrievals is critical for long-term retention, whether spacing is distributed in an equal or expanding manner is not as important (Karpicke & Bauernschmidt, 2011).

Another issue related to the review schedule is the effects of part and whole learning. Whole learning refers to a practice schedule where the target items are repeated as a whole. For example, when 100 words are studied in whole learning, 100 words are learned in sequence in one large deck of 100 items. Part learning, in contrast, refers to a practice schedule where the target items are repeated in smaller decks. In part learning, for instance, 100 words may be divided into five decks of 20 items and repeated in these smaller decks. Although part learning is often considered more effective than whole learning (e.g., Kornell, 2009; Wissman et al., 2012), research suggests that whole learning is actually more effective (Kornell, 2009). The results are possibly due to spacing. For instance, when 100 words are studied in whole learning, repetitions of a given target word are intervened by 99 other items. However, when 100 words are repeated in five decks of 20 items, repetitions of a given target word are intervened by only 19 other items. Whole learning, as a result, has a longer lag between repetitions (99 intervening items) than part learning (19 intervening items). Considering that a longer lag generally facilitates long-term retention more than a shorter lag (the lag effect), the advantage of whole learning in previous research may be due in part to differential lags rather than the part-whole distinction per se. This view is supported by Nakata and Webb (2016a), who compared the effects of part and whole learning while controlling spacing.

In Nakata and Webb (2016a), Japanese university students studied 20 English-Japanese word pairs using flash card software under part or whole learning conditions. Unlike previous studies, part and whole learning had equivalent lags between retrievals. Posttest results found no significant differences between the two conditions. In their Experiment 2, Nakata and Webb compared the effects of the following three treatments: (1) whole learning, (2) part learning with short lags, and (3) part learning with equivalent lags to whole learning. They found that (1) when part learning and whole learning have equivalent lags, there is no significant difference between the two schedules; (2) when part learning has shorter lags than whole learning, part learning is less effective than whole learning; and (3) part learning with longer lags is more effective than part learning with shorter lags. These results suggest that it is spacing, not the part-whole distinction, that affects learning.

Effects of Feedback

Another factor that is found to affect flash card learning is feedback. Feedback refers to the provision of the correct answer after retrieval. For instance, suppose learners were asked to translate *dog* into Japanese. If learners are provided with the correct answer (*inu*) after the retrieval attempt, the treatment involves feedback. Although retrieval without feedback can increase learning (e.g., Karpicke & Roediger, 2008), the provision of feedback enhances learning even more (e.g., Metcalfe & Kornell, 2007). One interesting issue regarding feedback is whether it should be provided immediately after the retrieval attempt (immediate feedback) or after a delay (delayed feedback). Some researchers argue that delayed feedback is more effective than immediate feedback because the former introduces longer lags between the retrieval attempt and feedback and according to the lag effect, longer lags often enhance learning more than shorter lags (e.g., Butler, Karpicke, & Roediger, 2007; Metcalfe, Kornell, & Finn, 2009).

Nakata (2015b), however, did not find any benefits of delaying feedback for L2 vocabulary learning. In his study, 98 Japanese university students studied 16 English-Japanese words pairs using flash card software. For half of the target items, the correct response was provided immediately after each retrieval attempt (immediate feedback). For the other half, the correct response was provided only after all 16 target items were practiced (delayed feedback). Nakata did not find any significant differences in immediate or delayed posttest scores between the two types of feedback. Considering that non-L2 vocabulary research has found benefits of delayed feedback (e.g., Butler et al., 2007; Metcalfe et al., 2009), it may be useful to further examine the effects of feedback timing on L2 vocabulary learning.

Effects of Interference

In flash card learning, some learners choose to study words that are semantically related to each other such as near-synonyms (e.g., *big*, *large*), antonyms (e.g., *hot*, *cold*), or coordinates (e.g., *cat*, *dog*). Researchers argue that presenting vocabulary in semantically related sets should be avoided because it decreases learning by causing interference (or cross-associations) among related words (e.g., Nation, 2000, 2013; Nation & Webb, 2011; Schmitt, 2010). However, empirical evidence is rather mixed regarding the effects of learning semantic sets. While some studies have shown the inhibitory effects of learning semantic sets (e.g., Tinkham, 1993, 1997; Waring, 1997), other studies have failed to do so (e.g., Hoshino, 2010; Ishii, 2015; Nakata & Suzuki, in press; Schneider et al., 2002). Ishii (2015), for instance, found that although physical similarity (e.g., *pencil*, *fishing pole*, *chopsticks*) inhibits learning, semantic similarity (e.g., *chicken*, *pig*, *giraffe*) may not. Nakata and Suzuki (in press) also compared the effects of learning semantically related and unrelated sets. One significant aspect of their study is that they ensured that the semantically related and unrelated sets were controlled for lexical variables that might affect learning other than semantic relatedness (e.g., word length, frequency, familiarity, imageability, pronounceability). Unlike some earlier research (e.g., Tinkham, 1993, 1997; Waring, 1997), they also gave a one-week delayed posttest to measure long-term retention. Nakata and Suzuki found that although semantic sets led to more cross-association errors than unrelated sets, no significant difference existed between the semantically related and unrelated sets in posttest scores. The results, together with those of other recent research (e.g., Hoshino, 2010; Ishii, 2015; Schneider et al., 2002), indicate that the negative effects of semantic sets might not be as robust as researchers have claimed.

Computer-Based Flash Cards

Recently, many flash card programs have become available as computer software, smartphone applications, or web-based applications (Nakata, 2011). Computer-based flash cards have several advantages over traditional paper-based flash cards. First, computer-based flash cards can help learners to implement principles of effective learning. Research suggests that many learners are unaware of principles for effective learning such as retrieval or spacing (e.g., Kornell, 2009; Nakata & Suzuki, in press; Wissman et al., 2012). Flash card software is useful because it can be programmed to ensure that learning follows principles for effective flash card learning. Second, computer-based flash cards can offer exercises that are hard to implement with paper-based flash cards. For instance, multiple-choice questions or exercises involving sounds (e.g., listening to the pronunciation of a word and typing the written form or its meaning) can be implemented relatively easily with computer-based flash cards. Third, flash card software can also keep record of learners' performance and determine the scheduling so that unfamiliar items are reviewed regularly after sufficient spacing (Nakata, 2011). Computers can also be programmed to manipulate not only scheduling but also retrieval formats. For instance, in earlier stages, flash card software can use relatively easy formats (e.g., multiple choice) and only in later stages introduce more challenging formats (e.g., recall of meaning or form). Gradually increasing difficulty is helpful according to the retrieval effort hypothesis, which states that difficult retrievals facilitate learning more than easy retrievals (Pyc & Rawson, 2009). Fourth, multimedia information, such as images, videos, and sounds, can be added to computer-based flash cards relatively easily. Fifth, some flash card programs facilitate flash card creation by supporting data entry. For instance, *Quizlet* (https://quizlet.com/), a popular web-based flash card program, automatically suggests the meanings and images corresponding to lexical items entered by users. It would also be helpful if flash card programs could automatically provide information such as the part of speech, pronunciation, collocations, contexts, frequency levels, inflected forms, derivations, or associations.

Limitations of Flash Card Learning

Although flash card learning is useful for vocabulary development, it is not without limitations. One limitation is that flash card learning typically focuses on initial form-meaning mapping and may not necessarily facilitate the learning of other aspects such as collocations, associations, or constraints on use (Nation, 2013). Another limitation is that it does not necessarily promote generative use. Generative use refers to a situation where learners encounter (at least partially) known words in new contexts (e.g., Joe, 1998). Imagine, for example, that a learner was first exposed to the English word *round* in the phrase "a round face". This learner might assume that the word *round* is an adjective that describes the shape of a concrete object. If this learner subsequently meets collocations such as *a round of drinks*, *a round of applause*, *a round number*, *a round dozen*, or *a round table*, s/he may be able to acquire a more precise understanding of this word. Flash card learning is not very effective in promoting generative use because target items are usually encountered without context or in a single context. Furthermore, flash card learning does not necessarily promote negotiation, instantiation, or imaging, all of which are found to facilitate vocabulary learning (Nakata & Webb, 2016b; Nation & Webb, 2011). Due to the aforementioned limitations, although flash card learning is one of the most effective vocabulary learning strategies, it should be supplemented with more meaning-focused activities, such as extensive reading, listening, or viewing.

Future Directions

Research on paired-associate learning dates back to the 19th century, and a number of studies have been conducted since. Nonetheless, there are still some issues related to flash card learning that warrant further investigation. First, considering that the duration of the treatment is relatively short in most existing studies, it is useful to conduct more longitudinal research on flash card learning. Second, because most existing studies on flash card learning define vocabulary acquisition as the ability to connect an L2 word with its meaning, research examining the effects of flash card learning on other aspects of vocabulary knowledge is warranted (for examples, see Webb, 2007, 2009a, 2009b). It will also be useful to examine the effects of flash card learning not only on explicit but also implicit vocabulary knowledge using priming (e.g., Elgort, 2011; Elgort & Piasecki, 2014) or event-related brain potential (Chun, Choi, & Kim, 2012).

One useful direction for future research may be to draw on findings from cognitive or educational psychology literature. Numerous studies have been conducted on paired-associate learning in not only the field of applied linguistics but also psychology. For instance, psychologists have examined the effects of a number of factors related to flash card learning such as spacing, retrieval, keyword mnemonics, and interference (for a review, see Dunlosky et al., 2013). Because most psychology studies have investigated the learning of materials other than L2 vocabulary, such as L1 vocabulary, some of the findings may not necessarily be applicable to L2 vocabulary learning. At the same time, they have nonetheless provided useful starting points for many L2 vocabulary studies (e.g., Barcroft, 2002, 2007; Nakata, 2015a, 2015b, 2016, 2017; Nakata & Webb, 2016a). It may be useful for applied linguists to empirically test the findings of the cognitive or educational psychology studies in the context of L2 vocabulary learning.

Another useful direction for future research is to examine whether different learners and lexical items benefit differently from flash card learning. In a grammar learning study, Suzuki and DeKeyser (2017) found that learners with different language aptitude (working memory capacity and language-analytic ability) benefit differently from massing and spacing. It would be useful to examine whether language learning aptitude affects flash card learning. For instance, do learners with higher working memory capacity benefit more from long spacing or whole learning? Another interesting issue is whether different lexical items benefit differently from flash card learning. Nakata and Suzuki (in press), for instance, found that although spacing facilitated the retention of both semantically related (e.g., *porcupine*, *raccoon*) and unrelated lexical sets (e.g., *berth*, *ointment*), semantically unrelated items benefited more from spacing. It would be useful to examine what kinds of words are particularly amenable to flash card learning. Since flash card learning is one of the most effective, efficient, and useful vocabulary learning strategies, future research on flash card learning is valuable for teachers, learners, and materials developers.

Further Reading

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Chapter 11 of this book is devoted entirely to flash card learning. It surveys a number of important studies on flash card learning and discusses principles for effective flash card learning.

Nakata, T., & Webb, S. (2016b). Evaluating the effectiveness of vocabulary learning activities using Technique Feature Analysis. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 123–138). New York, NY: Routledge.

This book chapter analyzes three common vocabulary learning activities including flash card learning. The chapter discusses advantages and disadvantages of flash card learning and proposes how to optimize L2 vocabulary learning from flash cards.

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

Although none of the chapters specifically addresses flash card learning, a number of issues related to flash card learning are discussed throughout the book. Part 3 of the book is particularly valuable for researchers planning to conduct experiments on flash card learning because it addresses a number of important considerations in designing experiments.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam: John Benjamins.

This book discusses a number of key concepts, such as retrieval, spacing, interference, mnemonics, levels (depth) of processing, transfer-appropriate processing, and Type of Processing – Resource Allocation (TOPRA) model, which have implications for effective flash card learning.

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.

This article proposes 17 criteria for evaluating computer-based flash card programs drawn from research on paired-associate learning. The article then analyzes nine freely or commercially available flash card software programs using these criteria.

Related Topics

Deliberate vocabulary learning, strategies for learning single-word items, resources for learning single-word items, evaluating exercises for learning vocabulary

References

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

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

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

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam, Netherlands: John Benjamins.

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

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. https://doi.org/10.1037/1076–898X.13.4.273

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 92–101. https://doi.org/10.1037/a0024828

Chun, E., Choi, S., & Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters*, *521*, 125–129. https://doi.org/10.1016/j.neulet.2012.05.069

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*, 4–58. https://doi.org/10.1177/1529100612453266

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Elgort, I., & Piasecki, A. E. (2014). The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition*, *17*, 572–588. https://doi.org/10.1017/S1366728913000588

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign-language vocabulary learning. *Language Learning*, *43*, 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

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

Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal*, *41*, 301–312. https://doi.org/10.1177/0033688210380558

Hulstijn, J. H. (1997). Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 203–224). Cambridge, UK: Cambridge University Press.

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.

Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research*, *19*, 712–722. https://doi.org/10.1177/1362168814559799

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*, 357–377. https://doi.org/10.1093/applin/19.3.357

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*, 1544–1550. https://doi.org/10.3758/s13423-014-0636-z

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. https://doi.org/10.1037/a0023436

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

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*, 1297–1317. https://doi.org/10.1002/acp.1537

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, *5*, 223–250. https://doi.org/10.1075/eurosla.5.11lau

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

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, *28*, 89–108. https://doi.org/10.1177/003368829702800106

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin and Review*, *14*, 225–229.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, *37*, 1077–1087. https://doi.org/10.3758/MC.37.8.1077

Morimoto, S., & Loewen, S. (2007). A comparison of the effects of image-schema-based instruction and translation-based instruction on the acquisition of L2 polysemous words. *Language Teaching Research*, *11*, 347–372.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022–5371(77)80016–9

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. https://doi.org/10.1080/09588221.2010.520675

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*, *37*, 677–711. https://doi.org/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. https://doi.org/10.1177/1362168814541721

Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, *54*, 257–289. https://doi.org/10.1515/iral-2015–0022

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*, 653–679. https://doi.org/10.1017/S0272263116000280

Nakata, T., & Suzuki, Y. (in press). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263118000219

Nakata, T., & Webb, S. (2016a). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, *38*, 523–552. https://doi.org/10.1017/S0272263115000236

Nakata, T., & Webb, S. (2016b). Evaluating the effectiveness of vocabulary learning activities using Technique Feature Analysis. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 123–138). New York, NY: Routledge.

Nation, I. S. P. (1980). Strategies for receptive vocabulary learning. *RELC Guidelines*, *3*, 18–23.

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

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

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

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

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *Modern Language Journal*, *80*, 478–493. https://doi.org/10.1111/j.1540-4781.1996.tb05468.x

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927. https://doi.org/10.3758/BF03192925

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Rodriguez, M., & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, *50*, 385–412. https://doi.org/10.1111/0023-8333.00121

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. https://doi.org/10.1002/acp.1083

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199–227). Cambridge, UK: Cambridge University Press.

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

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419–440. https://doi.org//10.1006/jmla.2001.2813

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. https://doi.org/10.1017/S0272263107070271

Suzuki, Y., & DeKeyser, R. (2017). Exploratory research on second language practice distribution: An Aptitude × Treatment interaction. *Applied Psycholinguistics*, *38*, 27–56. https://doi.org/10.1017/S0142716416000084

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

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*, 138–163. https://doi.org/10.1191/026765897672376469
Verspoor, M., & Lowie, W. (2003). Making sense of polysemous words. *Language Learning*, *53*, 547–586.
Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, *25*, 261–274. https://doi.org/10.1016/S0346-251X(97)00013-4
Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*, 63–81. https://doi.org/10.1177/1362168806072463
Webb, S. (2009a). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review*, *65*, 441–470. https://doi.org/10.1353/cml.0.0046
Webb, S. (2009b). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*, 360–376. https://doi.org/10.1177/0033688209343854
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*, 43–69. https://doi.org/10.1177/1362168814541734
Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, *20*, 568–579. https://doi.org/10.1080/09658211.2012.687052