What Spaced Learning is Effective for Long-Term L2 Vocabulary Retention?

Kohei KANAYAMA  
Sapporo Otani High School  
Kiwamu KASAHARA  
Hokkaido University of Education

Abstract

The present study aims to compare the effects of expanding, equally-spaced, and contracting retrieval practice on L2 vocabulary learning. Group A (n = 34) learned 20 English and Japanese word pairs (e.g., ligament-jintai) under the expanding schedule (Day 1, 1, 8, and 22). In a similar way, Group B (n = 19) and Group C (n = 29) learned the same target words under either the equally-spaced (Day 1, 8, 15, and 22), or the contracting schedule (Day 1, 15, 22, and 22), respectively. Twenty-one days after the last learning session (Day 43), all the groups took a delayed post-test in which they were required to retrieve Japanese meanings for the target words (e.g., ligament:______). The results showed that: (a) Group A had a better score than did Groups B and C after the second learning session; (b) Group C outperformed Groups A and B after the fourth learning session. (c) However, no significant difference was observed between the three groups in the delayed test. These findings indicate that the length of the interval between each vocabulary session does not matter for long-term L2 vocabulary retention as long as learners are given three or more spaced learning sessions.

1. Introduction

1.1 The Importance of Linking Between the L2 Form and its L1 Translation Connection

A large number of studies have indicated that L2 learners should encounter new words in meaningful contexts or have a large amount of reading and listening input (Ellis, 2002) because this is almost the same process as when L1 learners acquire new words in their native languages. On the other hand, many scholars and language teachers have made the criticism that paired-associate learning, where learners intentionally focus on making connections between an L2 form and its L1 meaning, is not useful because it is not a natural process of language learning and does not contribute to normal language use (Oxford & Crookall, 1990). That is, making form-and-meaning connections does not include learning the functions of the target words, which is indispensable to real communication. Moreover, this criticism has been strongly supported by the acquisition-learning
theory (Krashen, 1985), which proposes that obtaining L2 knowledge subconsciously, which is identical to what babies go through in their L1, (acquisition) can lead to real L2 communication, but acquiring L2 knowledge intentionally or consciously (learning) is useless in actual L2 language use. Thus, a significant number of communicative language teaching proponents have suggested that words obtained with paired-associate learning may not be useful for real L2 communication.

However, recent studies argue that deliberately learned vocabulary knowledge is essential for second language learning for the following two reasons. First, even intentionally learned vocabulary knowledge can be accessed subconsciously and fluently, and can be available for normal language use (Elgort, 2011; Nation, 2013). The study of Elgort demonstrated that deliberately learned word knowledge was automatically integrated into usable knowledge. Her study discredited the assumption that intentional vocabulary learning is not helpful for language use. Rather, as Nation (2013) insists, connecting the form to the meaning is very useful prerequisites to using a word. Second, it is almost impossible for learners in an English as a foreign language (EFL) environment to encounter new words only through natural meaningful input or contexts. Kanayama and Kasahara (2015b) argue that it is possible to acquire a good deal of target vocabulary through meaningful input in an English as a second language (ESL) situation, where a sufficient amount of input is guaranteed. However, there are few opportunities to obtain English input outside the classroom in an EFL environment. Paired-associate learning is a crucial first step for EFL learners to remember the meanings of a large number of new words within a short period of time.

In sum, ideally, learners should establish strong links between L2 forms and their L1 translations as a first step in L2 vocabulary learning (Folse, 2004). After that, they are able to gain great benefits from incidental learning and to increase other aspects of vocabulary knowledge such as grammatical function, collocation, and constraints on use. Therefore, the present study aims to show how to enhance form-and-meaning connections deliberately at an early stage of L2 vocabulary learning.

1.2 The Theoretical Reasons for Spaced Retrieval Practice over Massed Practice

Vocabulary learning requires learners to have opportunities to encounter the same target words repeatedly (Webb, 2007; Nation, 2013). Just one study trial does not guarantee successful vocabulary learning, even if efficient methods are used (Kanayama & Kasahara, 2016). Additionally, a considerable number of studies have shown that learning is enhanced when study sessions are spaced apart, rather than occurring close together (Pyc & Rawson, 2007; Bell, Kawadri, Simone, & Wiseheart, 2014). This is known as distributed practice or spaced retrieval practice (Kornell, 2009; Baddeley, 2013; Kapler, Weston, & Wiseheart, 2015).

The main reason why spaced retrieval practice is more effective in long-term retention than massed practice is associated with the principle of retrieval effort. Effective learning occurs if learners make more effort in retrieving an item from memory (Küpper-Tetzel, Kapler, & Wiseheart, 2014; Kanayama & Kasahara, 2015b). Usually, learners using spaced retrieval practice gain greater
benefit from retrieval effort than those using massed practice. Distributed practice requires learners to make more effort to retrieve an item during the recall session because of the intervals between the learning sessions. During an interval between each study session, learners forget some of the learned information, which requires the learners to make more effort when recalling the learned information in the following learning session. This attempt to recall the information from memory enhances long-term retention of the learned information (Roediger & Karpicke, 2006a, 2006b; Roediger, Agarwal, McDaniel, & McDermott, 2011). In sum, spaced retrieval practice can ensure learners make more effort, which leads to better long-term retention (Logan & Balota, 2008; Karpicke & Roediger, 2010; Nakata, 2015). On the other hand, however, learners using massed practice cannot benefit from retrieval effort. Massed practice involves repetitions over short periods of time (Kornell, 2009), and they can review items before forgetting occurs. They are able to constantly retrieve items successfully from the beginning. Such a successful experience without much retrieval effort provides them with an impression of mastering everything, and then the learners gradually make less effort in recalling the items afterward.

Another reason for inferiority of massed practice to spaced retrieval practice is that items learned in cramming tend to be forgotten soon afterwards (Baddeley, 2013). One of the most powerful supporters of this idea was Hermann Ebbinghaus, who reported that learners forgot 74% of the learned items just one day later (Ebbinghaus 1885/1964). This indicates that cramming study sessions together cannot contribute greatly to long-term retention, even if repeating an item several times is involved (Kanayama & Kasahara, 2016). These are the reasons why spaced retrieval practice is more effective in long-term retention than massed practice. This is also the case with L2 vocabulary learning. Learners should have several vocabulary learning sessions with a certain time interval between each session (Sobel, Cepeda, & Kapler, 2011).

1.3 The Effective Interval Between Each Vocabulary Session

Although several studies have shown the clear advantage of spaced retrieval practice, it remained to be seen how far each study session should be spaced apart. Indeed, there are three types of relative spacing: expanding retrieval practice, equally-spaced retrieval practice, and contracting retrieval practice. Relative spacing refers to how far apart study sessions are spaced from one another (Karpicke & Bauernschmidt, 2011; Nakata, 2015). Expanding retrieval practice requires students to relearn an item immediately after learning it initially, and then gradually increases the interval between each study session (Karpicke & Roediger, 2007, 2010; Nation, 2013; Kanayama & Kasahara, 2015a). On the other hand, equally-spaced retrieval practice means that the intervals between each learning trial are kept constant (Logan & Balota, 2008; Nakata, 2015). Additionally, in the contracting retrieval practice, the interval between each learning trial is gradually decreased across time (Küpper-Tetzel, et al., 2014).

Traditionally, expanding retrieval practice has been thought to be more effective in long-term retention than the other two types of relative spacing (Nakata, 2015) because learners using
expanding retrieval practice can gain greater benefits from both retrieval effort and a successful retrieval experience. Learning is promoted if learners use more effort in retrieving an item from memory and then again if the retrieval is successful (Pyc & Rawson, 2009; Karpicke, Lehman, & Aue, 2014; Nakata, 2015). Expanding retrieval practice can keep the probability of a high successful retrieval rate (Landauer & Bjork, 1978). Basically, learners using expanding retrieval practice have a chance to undertake the first retrieval session soon after the first study session. Thus, they have a great chance to recall an item successfully, before too much forgetting occurs (Nation, 2013). Moreover, the remaining retrieval sessions require the learners to keep up efforts to retrieve the item because the interval between each session is gradually longer. In other words, expanding retrieval practice can ensure a successful retrieval experience (Karpicke & Roediger, 2010; Storm, Bjork, & Storm, 2010; Karpicke & Bauermischmidt, 2011; Küpper-Tetzel et al., 2014) and a constant retrieval rate (Landauer & Bjork, 1978). Basically, learners using expanding retrieval practice have a great chance to recall an item successfully, before too much forgetting occurs (Nation, 2013).

1.4 Comparison Between Expanding and Equally-Spaced Retrieval Practice

However, recent studies have found no clear advantages of expanding practice over the equally-spaced retrieval practice (Cull, 2000; Kang, Lindsey, Mozer, Pashler, 2014; Kanayama & Kasahara, 2015a, 2016). For example, in a study of Kang et al. (2014), the expanding group learned 20 Japanese-English word pairs on Days 1, 3, 9, and 28, while the equally-spaced group learned another 20 Japanese-English pairs on Days 1, 10, 19, and 28. Additionally, a delayed post-test was carried out 56 days after the last learning session (Day 84). No significant difference was found between the expanding and equally-spaced groups in the immediate test conducted on Day 28 (62% vs. 65%). Moreover, both groups had the same results in the delayed post-test conducted on Day 84 (expanding = 49%, equally-spaced = 46%).

The same results were also observed in our previous study (Kanayama & Kasahara, 2016). The expanding group (n = 34) remembered 20 English and Japanese word pairs (e.g., ligament-jimnntai), taking the four study sessions on Days 1, 1, 8, and 22. This means that Day 1 had two study sessions with an hour interval between them. On the other hand, participants in the equally-spaced group (n = 19) learned the same 20 words on Days 1, 8, 15, and 22. Both groups took immediate tests after each 5-minute study session. Moreover, they also took the delayed post-test 21 days after the last learning session (Day 43). All the tests asked them to recall Japanese translations for the target English words (e.g., ligament:______?).

The results found that the expanding group had a better score than did the equally-spaced group in the second immediate test (70.3% vs. 51.3%). However, the significant difference disappeared in all the remaining tests. There was no significant difference between the expanding and equally-spaced group in the fourth immediate test (82.7% vs. 79.3%) or in the delayed test (58.1% vs. 52.6%). Therefore, based on these previous studies, it was concluded that the number of repetitions is more important for enhancing the linkage between L2 forms and their L1 equivalents than length of spacing. If learners are given three or more study sessions, relative spacing does not
expanding retrieval practice can gain greater benefits from both retrieval effort and a successful retrieval experience (Aue, 2014; Nakata, 2015). Expanding retrieval practice can keep the probability of a high successful retrieval rate high (Storm, 2010; Karpicke & Bauernschmidt, 2011; Küpper-Tetzel, et al., 2014) and a constant retrieval interval between each session is gradually longer. In other words, expanding retrieval practice can help the learners make a retrieval effort, develop effective learning plans from the experience of retrieval failure, and have future successful retrieval experiences.

Moreover, the remaining retrieval sessions require the learners to keep up efforts to retrieve the item successfully, before too much forgetting occurs (Nation, 2013). Therefore, based on these previous studies, it was concluded that the number of repetitions is more important for enhancing the linkage between L2 forms and their L1 equivalents (58.1% vs. 52.6%).

1.5 Potential Benefits of Contracting Retrieval Practice

One essential problem in previous studies (Cull, 2000; Dobson, 2012; Kang et al., 2014; Kanayama & Kasahara, 2015a; 2016) is that they none examined the effectiveness of contracting retrieval practice. Küpper-Tetzel et al. (2014) indicate the advantage of contracting retrieval practice for the following reason. It requires learners to make more effort while retrieving items than is the case for expanding and equally-spaced retrieval practice. Because learners using contracting retrieval practice have no chance to take the first retrieval session soon after the first study session, they have a high probability of retrieval failure there. However, the retrieval failure helps the learners develop an effective learning plan (Roediger, Putnam, & Smith, 2011; Kanayama & Kasahara, 2015b). The learners can distinguish the items they have mastered and the items they have not, and can make more effort in remembering the items in the remaining retrieval sessions (Son & Kornell, 2008). Finally, the learners have a great chance to recall the items successfully afterwards because the interval between each learning session is gradually decreased. In sum, contracting retrieval practice can help the learners make a retrieval effort, develop effective learning plans from the experience of retrieval failure, and have future successful retrieval experiences.

The only study which employed contracting retrieval practice was conducted by Küpper-Tetzel, et al. (2014), as far as we found. Two hundred and ten university students were asked to remember 28 cue-target word pairs under either expanding (Day 1, 2, and 7), equally-spaced (Day 1, 4, and 7), or contracting schedules (Day 1, 6, and 7). After all the learning sessions, all the groups took a delayed post-test in which they recalled the target words for the cue words. The post-test was carried out for 15 minutes, 1 day, 7 days, and 35 days after the last learning session, respectively. In the 1-day and 7-day delayed test, the participants with the contracting schedule had a significantly higher score than did the other groups, whereas no significant difference was observed between the expanding and equally-spaced group. Surprisingly, the results were reversed in the 35-day delayed test. The expanding and equally-spaced groups had significantly better scores than did those with the contracting schedule. Therefore, this study did not show the advantages of contracting retrieval practice over the other two types of relative spacing for longer retention.

1.6 Purpose of the Present Study

Most of the previous studies have focused on the comparison between expanding and equally-spaced retrieval practice. They did not investigate the potential benefits of contracting retrieval practice. Our previous studies (Kanayama & Kasahara, 2015a, 2016) also compared the effectiveness of only two kinds of relative spacing, expanding and equally-spaced retrieval practice. The only study that dealt with contracting retrieval practice (Küpper-Tetzel, et al., 2014) did not prove that it was more effective for longer retention. A possible reason could be that the period of the whole learning sessions was only a week, from Day 1 to Day 7. For long-term L2 vocabulary
retention, a longer-term experiment should have been conducted. Therefore, it has remained to be seen which is the most effective form of relative spacing in long-term retention. The present study aims to compare the effects of three forms of relative spacing, with a 22-day learning session. This study combines the data of the expanding and equally-spaced groups obtained in Kanayama and Kasahara (2016), with the data of the contracting retrieval group obtained in this experiment. Hence, the present study employs the same procedure, materials (20 English and Japanese word pairs), and pre-test (see Appendix B) as used in the study of Kanayama and Kasahara (2016). The research question of the present study is as follows:

RQ. Which types of relative spacing are the most effective for long-term L2 vocabulary learning?

2. Method

2.1 Participants

The participants were 82 first-grade Japanese EFL university students who belonged to three general English classes. All of the learners had studied English for a minimum of six years. The expanding retrieval group (Day 1, 1, 8 and 22) was designated to Group A (n = 34), the equally-spaced retrieval group (Day 1, 8, 15 and 22) to Group B (n = 19), and the contracting retrieval group (Day 1, 15, 22, and 22) to Group C (n = 29). In order to discover if there was any difference in vocabulary size between the groups, the Vocabulary Size Test (Nation & Beglar, 2007) was implemented for the three groups at the 1000, 2000 and 3000 word levels (Maximum score = 30). The mean score of Group A (M = 19.91, SD = 2.41) was quite similar to that of Group B (M = 19.84, SD = 3.07), and Group C (M = 19.75, SD = 3.09). A one-way ANOVA found no significant difference between the three groups, F (2, 79) = .023, p = .977. This confirmed that there was no significant difference in vocabulary size among the groups.

2.2 Materials

The present study employed the same L2 low-frequency English words and L1 Japanese equivalent pairs that had been used in our previous studies (Kanayama & Kasahara, 2016). Before the experimental session, 50 low-frequency words including the 20 target words, which are not listed in JACET 8000 (JACET, 2003), were presented to the contracting retrieval group as a pre-test (see Appendix B). The participants in Group C had to say if they knew any of the 50 words. We confirmed that none of the participants in Group C knew any of the target 20 words used in Kanayama and Kasahara (2016). All the 10 nouns and 10 verbs are shown in Table 1 with their Japanese translations.

2.3 Procedures

All the learning sessions and the tests were conducted in general English lessons regularly.
held at the university. Four learning sessions were given to each group due to the number of lessons available for the experiment. On Day 1, all the participants were given the same word list as used in Kanayama and Kasahara (2016), with the 20 English target words along with their Japanese translations (see Appendix A). In case a participant preferred to memorize the target words by writing them, a blank sheet of paper was also given out. After a 5-minute vocabulary learning session, the first immediate test session took place, where the participants were required to write each word’s meaning in Japanese within five minutes (e.g., ligament:____?). They took part in the same learning session and the immediate recall test in all the remaining learning phases.

For the participants in Group A (expanding group), the second learning session occurred only an hour after the first learning session (Day 1). The third learning session also was run seven days after the second learning session (Day 8). Fourteen days after the third learning session, the participants attended the fourth learning session (Day 22). Therefore, the expanding group took part in all the four learning sessions on Days 1, 1, 8, and 22. The learning schedules for Group B (equally-spaced group) remained the same (Day 1, 8, 15, and 22).

Each interval period was gradually decreased for the participants in Group C (the contracting group). The second learning session took place two weeks after the first learning session (Day 15). A week after the second learning session, the participants attended the third learning session (Day 22). Moreover, an hour after the third learning session, the fourth learning session was conducted (Day 22). Finally, all the groups took the delayed post-test 21 days after the fourth learning session without the 5-minute learning session (Day 43) and with no advance notice. Table 2 presents the schedule for each learning session, the immediate test and the delayed post-test.

<table>
<thead>
<tr>
<th>Nouns</th>
<th>Equivalent</th>
<th>Verbs</th>
<th>Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>gnome</td>
<td>格言</td>
<td>outwit</td>
<td>(人を)出し抜く</td>
</tr>
<tr>
<td>mutineer</td>
<td>反逆者</td>
<td>gnaw</td>
<td>(ネコを)を〜かじる</td>
</tr>
<tr>
<td>ointment</td>
<td>化粧用クリーム</td>
<td>exasperate</td>
<td>(人を)怒らせる</td>
</tr>
<tr>
<td>adhesive</td>
<td>接着剤</td>
<td>nauseate</td>
<td>嫌悪を感じさせる</td>
</tr>
<tr>
<td>lemur</td>
<td>キツネザル</td>
<td>solicit</td>
<td>懇願する</td>
</tr>
<tr>
<td>portent</td>
<td>前兆</td>
<td>bemoan</td>
<td>〜を嘆き悲しむ</td>
</tr>
<tr>
<td>ligament</td>
<td>鞄帯</td>
<td>smirk</td>
<td>にやにや笑う</td>
</tr>
<tr>
<td>hoof</td>
<td>(馬の)ひづめ</td>
<td>smuggle</td>
<td>〜を密輸する</td>
</tr>
<tr>
<td>encroachment</td>
<td>侵略</td>
<td>sterilize</td>
<td>〜を殺菌する</td>
</tr>
<tr>
<td>palliative</td>
<td>緩和剤</td>
<td>admonish</td>
<td>心告する</td>
</tr>
</tbody>
</table>
2.4 Scoring

One of the authors graded students’ score. Two points were given for each correct answer. However, one point was given for a partially correct answer if (a) an intransitive verb was given for a transitive verb (e.g. *okoru* for “exasperate”) and vice versa, or (b) participants made a mistake concerning a part of speech (e.g. *shinryaku-suru* for “encroachment”). The maximum score for each test was 40 points (2 points × 20 words).

2.5 Data Analysis

A 3 (Group: Group A, Group B, Group C) × 5 (Test: Immediate Test 1, Immediate Test 2, Immediate Test 3, Immediate Test 4, Delayed Test 1) mixed ANOVA was conducted in order to examine the research question. In addition, a simple main effect test was carried out between the mean scores of Group A, Group B, and Group C in each immediate and delayed test, respectively.

Table 2

<table>
<thead>
<tr>
<th>Schedules for Learning Sessions, Immediate Tests, and the Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; learning &amp; test session</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; interval</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; learning &amp; test session</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; interval</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; learning &amp; test session</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; interval</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; learning &amp; test session</td>
</tr>
<tr>
<td>21-days interval</td>
</tr>
</tbody>
</table>

3. Results

Table 3 shows the means and the standard deviations of each immediate test and the delayed post-test for Group A, Group B, and Group C. Figure 1 presents the changes of the means for each immediate and delayed post-test for Group A, Group B, and Group C. A 3 (Group: Group A, Group B, Group C) × 5 (Test: Immediate Test 1, Immediate Test 2, Immediate Test 3, Immediate Test 4, Delayed Test 1) mixed ANOVA found that there was a significant interaction between the two factors, Group and Test, $F (2, 79) = 6.32$, $p < .001$, $\eta^2 = .13$ (large effect size). Additionally, there was a main effect of Test, $F (2, 79) = 230.68$, $p <.001$, $\eta^2 = .74$ (large effect size), but no main effect of Group was found, $F (2, 79) = 1.76$, $p = .177$, $\eta^2 = .04$. 

Furthermore, there was a simple main effect between the three groups in Immediate Test 2, \( F(2, 79) = 7.09, p < .001 \), and in Immediate Test 4, \( F(2, 79) = 4.56, p = .01 \). However, no simple main effect was found between the three groups in Immediate Test 1, \( F(2, 79) = 1.11, p = .329 \), in Immediate Test 3, \( F(2, 79) = 1.25, p = .286 \), or in Delayed Test 1, \( F(2, 79) = .63, p = .50 \).

**Table 3**

*Means, Standard Deviations of each Immediate Test and Delayed Post-Test for Group A, Group B, and Group C (\( N = 82 \), Full score = 40).*

<table>
<thead>
<tr>
<th>Test (Expanding)</th>
<th>Group A Mean</th>
<th>SD</th>
<th>Group B Mean</th>
<th>SD</th>
<th>Group C Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate Test 1</td>
<td>1 16.35  7.76</td>
<td>1 13.37  5.89</td>
<td>1 15.75  6.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate Test 2</td>
<td>1 28.15  8.03</td>
<td>8 20.53  8.56</td>
<td>15 22.31  7.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate Test 3</td>
<td>8 30.32  8.11</td>
<td>15 28.68  8.92</td>
<td>22 32.03  6.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate Test 4</td>
<td>22 33.09  7.48</td>
<td>22 31.74  8.21</td>
<td>22 37.82  3.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed Test 1</td>
<td>43 23.24  8.8</td>
<td>43 21.05  8.58</td>
<td>43 21.1  7.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** The means of each immediate and delayed post-test for Group A, Group B, and Group C.

Finally, Post-hoc multiple comparisons were carried out between the three groups in Immediate Test 2 and Immediate Test 4. There was a significant difference between the expanding and equally-spaced conditions in Immediate Test 2, \( t(51) = 3.50, p < .001, r = .44 \) (large effect size), and between the expanding and contracting conditions in Immediate Test 2, \( t(61) = 3.041, p = .002, r = .36 \) (medium effect size), but not between the equally-spaced and contracting conditions in Immediate Test 2, \( t(46) = .79, p = .426, r = .11 \) (small effect size). This means that the expanding group had a better score than the equally-spaced and the contracting retrieval groups in Immediate Test 2 (70.3% vs. 51.3% vs. 55.7%). In Immediate Test 4, there was a significant difference between the contracting and equally-spaced conditions, \( t(46) = 2.71, p = .006, r = .37 \) (medium effect size) and between the contracting and expanding conditions, \( t(61) = 2.46, p = .013, r = .30 \) (medium effect size).
equally-spaced and the contracting groups caught up with that of the expanding group after the third equally-spaced, and contracting groups (75.8% vs. 71.7% vs. 80%). The average scores of the 4.2 Results of Immediate Test 3 and 4 (the Later Stage of the Vocabulary Session)

which is why there was no significant difference in the results of Immediate Test 2 between them. contracting group. Both of the groups had to restart the learning session from almost the same level, and had to learn the target items from scratch in the second learning session, like the learners in the group for seven days. This means that the learners in the equally-spaced group lost their advantage forgetting occurs. It is not surprising that the expanding group got a significantly higher score than the contracting group, which had the second learning session 14 days after the first session. This result is consistent with the results of Kanayama and Kasahara (2015a; 2016).

However, no significant difference was found between the equally-spaced and the contracting groups in Immediate Test 2 (51.3% vs. 55.7%). If learners gained a great benefit from the second learning session soon after the first learning session, the equally-spaced group would have had a better score than did the contracting group because the interval between the initial learning session and the second learning session for the equally-spaced group was shorter than that for the contracting group (7 days vs. 14 days). A possible reason why there was no difference between the two groups in Immediate Test 2 is that both groups had forgotten most of the target words before the second learning session. Fourteen days was long enough for the participants in the contracting group to forget most of the items. This may have happened to the learners in the equally-spaced group for seven days. This means that the learners in the equally-spaced group lost their advantage and had to learn the target items from scratch in the second learning session, like the learners in the contracting group. Both of the groups had to restart the learning session from almost the same level, which is why there was no significant difference in the results of Immediate Test 2 between them.

4.3 Results of the Delayed Post-Test

Immediate Test 3 found that there was no significant difference between the expanding, equally-spaced, and contracting groups (75.8% vs. 71.7% vs. 80%). The average scores of the equally-spaced and the contracting groups caught up with that of the expanding group after the third learning session. Indeed, both the equally-spaced and the contracting groups had greater benefit effect size), but not between the expanding and equally-spaced conditions, \( t (51) = .62, p = .534, r = .08 \) (small effect size). This means that the contracting group had a better score than did the expanding and equally-spaced retrieval practice groups in Immediate Test 4 (94.5% vs. 82.7% vs. 79.3%).

4. Discussion

4.1 Results of Immediate Test 1 and 2 (the Early Stage of the Vocabulary Session)

Immediate Test 1 found that there was no significant difference between the three groups. This was a natural result because all the groups took the same procedure on Day 1. They were all asked to remember the target words in five minutes and to recall Japanese meanings for the target words in five minutes. However, Immediate Test 2 found that the expanding group showed higher performance than did the equally-spaced group and the contracting group in Immediate Test 2 (70.3% vs. 51.3% vs. 55.7%). Kanayama and Kasahara (2015a; 2016) have found that the second learning session carried out immediately after the first learning session can ensure learners’ successful retrieval because the learners have a chance to see the target words again before much forgetting occurs. It is not surprising that the expanding group got a significantly higher score than the contracting group, which had the second learning session 14 days after the first session. This result is consistent with the results of Kanayama and Kasahara (2015a, 2016).

However, no significant difference was found between the equally-spaced and the contracting groups in Immediate Test 2 (51.3% vs. 55.7%). If learners gained a great benefit from the second learning session soon after the first learning session, the equally-spaced group would have had a better score than did the contracting group because the interval between the initial learning session and the second learning session for the equally-spaced group was shorter than that for the contracting group (7 days vs. 14 days). A possible reason why there was no difference between the two groups in Immediate Test 2 is that both groups had forgotten most of the target words before the second learning session. Fourteen days was long enough for the participants in the contracting group to forget most of the items. This may have happened to the learners in the equally-spaced group for seven days. This means that the learners in the equally-spaced group lost their advantage and had to learn the target items from scratch in the second learning session, like the learners in the contracting group. Both of the groups had to restart the learning session from almost the same level, which is why there was no significant difference in the results of Immediate Test 2 between them.

4.2 Results of Immediate Test 3 and 4 (the Later Stage of the Vocabulary Session)

Immediate Test 3 found that there was no significant difference between the expanding, equally-spaced, and contracting groups (75.8% vs. 71.7% vs. 80%). The average scores of the equally-spaced and the contracting groups caught up with that of the expanding group after the third learning session. Indeed, both the equally-spaced and the contracting groups had greater benefit
from the third learning session than the expanding group, because the score of the expanding, the equally-spaced, and the contracting groups increased by 5.5%, 20.4%, and 24.3%, respectively from the previous test. The expanding group had already achieved a higher score in Immediate Test 2 (70.3%). On the other hand, both the equally-spaced and the contracting group achieved around 50 percent of correct answers (51.3% and 55.7%). Both groups had more room for improvement, and gained greater benefits from the third learning session than did the expanding group.

Immediate Test 4 showed that the contracting group had a better score than did the expanding and the equally-spaced groups (94.5% vs. 82.7% vs. 79.3%). Indeed, the score of the contracting, expanding, and the equally-spaced groups increased by 14.5%, 6.9%, 7.6%, respectively from the previous test. Why did the contracting group recall significantly more target words than the other two groups? This was due to the fact that the contracting group took the third and fourth learning session on the same day (Day 22). Our previous studies (Kanayama & Kasahara, 2015a; 2016) observed that the first review session immediately after the first learning session helped boost the scores in a recall test. Our previous and present studies have revealed that any review sessions soon after the previous learning session can temporally strengthen connections between L2 forms and their L1 equivalents.

### 4.3 Results of the Delayed Post-Test

Delayed Test 1 revealed that no significant difference was found between the expanding, equally-spaced, and contracting groups (58.1% vs. 52.6% vs. 52.7%), even though the contracting group outperformed the other two groups in Immediate Test 4. Therefore, this study reconfirmed the previous studies (Kang et al., 2014, Kanayama & Kasahara, 2015a, 2016), and revealed that any types of relative spacing can be similarly effective for long-term L2 vocabulary retention, as long as three or more learning sessions are guaranteed. The present study mainly focused on the effectiveness of contracting retrieval practice, but its superiority for long-term retention was not observed. Performance during each learning phase may not necessarily be a better index of long-term retention (Nakata, 2015).

The research question of this present study was to examine the effective relative spacing in long-term L2 vocabulary retention. The answer is that if learning sessions are distributed over multiple days, with three or more learning sessions, the length of time between the vocabulary sessions does not significantly affect the subjects’ total benefit from the learning. As Kanayama and Kasahara (2016) have indicated, it is the number of repetitions, not the length of spacing that is a decisive factor in enhancing long-term retention of word connections between L2 and L1 pairs. Learners can consolidate the linking between L2 target words and their L1 equivalents by repeated paired-associate learning. This study observed that any review sessions soon after the previous learning session improved the results of the students’ recall test. However, the types of relative spacing did not make difference in long-term retention. Hence, it is the number of repetitions over days that is the crucial factor in enhancing long-term retention of L2 and L1 connections. Teachers
and learners do not have to be overly concerned about which types of spaced retrieval practice contribute to better long-term retention as long as they know that learning is much enhanced when multiple study sessions are distributed over time periods (Carpenter et al., 2012). Based on the findings of the present study, there is a pedagogic implication. English teachers should space the vocabulary sessions apart. It is important that they give their students repeated instances of the same words throughout a given time period including several intervals.

Finally, a limitation of the present study should be noted. Neither the expanding nor the contracting groups took part in vocabulary sessions in a truly gradual learning schedule. The four learning sessions of the expanding group were arranged in the following schedule: Day 1, an hour later, Day 8 and Day 22. Similarly, the learning schedule for the contracting group was put on Day 1, Day 15, Day 22 and an hour later. The spacing between each vocabulary session was not gradual, but changed suddenly. The expanding and contracting retrieval practices in this study should have involved more gradually changing intervals between the sessions (Karpicke & Roediger, 2007, 2010; Nakata, 2015).

5. Conclusions

The purpose of the present study was to investigate the optimal interval between each vocabulary session, employing the potential benefits of contracting retrieval practice. Thus, the research question of the present study was to investigate which types of relative spacing would be the most effective for long-term retention of L2 vocabulary retention. This experiment found that the most crucial factor contributing to strong L2 and L1 connections is the number of repetitions with multiple spaced retrievals, but not the types of relative spacing (the length of the interval between each vocabulary session). Thus, as long as learning sessions are spaced apart, and include three or more learning sessions, the types of relative spacing do not matter for long-term L2 vocabulary retention.

References

Dobson, J. L. (2012). Effect of uniform versus expanding retrieval practice on the recall of
three or more learning sessions, the types of relative spacing do not matter for long-term L2 between each vocabulary session). Thus, as long as learning sessions are spaced apart, and include with multiple spaced retrievals, but not the types of relative spacing (the length of the interval the most crucial factor contributing to strong L2 and L1 connections is the number of repetitions the most effective for long-term retention of L2 vocabulary retention. This experiment found that the research question of the present study was to investigate which types of relative spacing would be involved more gradually changing intervals between the sessions (Karpicke & Roediger, 2007, Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and vocabulary retention.


physiology information. Advance in Physiology Education, 36, 6-12.


JACET. (2003). JACET List of 8000 Basic Words: JACET 8000. Tokyo; JACET.


**Appendix A: Learning Sheet (For Group A, Group B, and Group C)**

outwit: (人など)を出し抜く、裏をかく

gnaw: (ネコ・ネズミなどが)〜をかじる

gnome: 格言

exasperate: (人)を怒らせる

mutineer: 反逆者

nauseate: (人)に嫌悪を感じさせる

ointment: 化粧用クリーム

solicit: 懇願(こんがん: ひたすらお願いすること)する

adhesive: 接着剤

bemoan: 〜を嘆き悲しむ

lemur: キツネザル

portent: (不吉・重大事の)前兆、前触れ

smirk: にやにや笑う

smuggle: 〜を密輸する

sterilize: 〜を殺菌する、〜を消毒する

ligament: 韻帯(じんたい)
hoof: (馬などの)ひつじ

encroachment: 侵略(しんりゃく)

palliative: 緩和剤

admonish: 忠告する、警告する

**Appendix B: Pre-test (for Group A, Group B, and Group C)**

Q. あなたが知っている英単語に○を、知っているかもしれない英単語に△を、知らない単語には×を記入してください。

| 1. outwit [ ] | 18. teal [ ] | 35. jollity [ ] |
| 2. dissident [ ] | 19. yelp [ ] | 36. lemur [ ] |
| 3. encroachment [ ] | 20. yolk [ ] | 37. lisp [ ] |
| 4. gnaw [ ] | 21. zest [ ] | 38. nuke [ ] |
| 5. dissipate [ ] | 22. adhesive [ ] | 39. opacity [ ] |
| 6. gnome [ ] | 23. admonish [ ] | 40. pillage [ ] |
| 7. ideogram [ ] | 24. bemoan [ ] | 41. ply [ ] |
| 8. laze [ ] | 25. canary [ ] | 42. portent [ ] |
| 9. exasperate [ ] | 26. gnash [ ] | 43. scour [ ] |
| 10. mutineer [ ] | 27. casket [ ] | 44. smirk [ ] |
| 11. natter [ ] | 28. collude [ ] | 45. smuggle [ ] |
| 12. nauseate [ ] | 29. edifice [ ] | 46. sterilize [ ] |
| 13. ointment [ ] | 30. efface [ ] | 47. tote [ ] |
| 14. outage [ ] | 31. fatality [ ] | 48. insomnia [ ] |
| 15. palliative [ ] | 32. flit [ ] | 49. ligament [ ] |
| 16. rower [ ] | 33. hoof [ ] | 50. eradicate [ ] |
| 17. solicit [ ] | 34. immerse [ ] |