The Effects of Expanding and Equally-Spaced Retrieval Practice on Long-Term L2 Vocabulary Retention

Kohei KANAYAMA
Graduate School, Hokkaido University of Education
Kiwamu KASAHARA
Hokkaido University of Education

Abstract

The present study aims to compare the effects of expanding and equally-spaced retrieval practice on long-term L2 vocabulary retention. Since the discovery of Ebbinghaus’ forgetting curve, expanding retrieval practice has been considered more effective than equally-spaced retrieval practice. Learners have been encouraged to review target items immediately after the first learning session, and then extend intervals between study sessions gradually, rather than have the same interval period between each session. The present study challenges this assumption. Participants in Group A (n = 34) learned 20 target words under the expanding condition (Day 1, 1, 8 and 22), while those in Group B (n = 19) learned the same 20 target words under the equally-spaced condition (Day 1, 8, 15 and 22). Twenty-one days after the learning session (Day 43), both groups took a delayed post-test, where they were asked to recall Japanese meanings for the English target words. This study revealed that the expanding group showed significantly better results than the equally-spaced group just after the first review session. However, there was no significant difference in long-term retention between the two forms of spaced learning, as long as both groups were given the chance to have four learning sessions.

1. Introduction

Vocabulary acquisition is an essential part of second language learning. To date, a number of studies have focused on how English teachers present new words effectively to L2 learners, such as in collocations (Ishizuka, 2005; Kasahara, 2010; 2015), informative sentences (Webb, 2007a; 2008), with glosses (Kasahara, 2005; Webb, 2007b), paired-associate learning (Prince, 1996; Laufer & Shmueli, 1997) and on word cards (Nakata, 2008).

However, it is not possible for students to acquire the new words completely in just one learning session, even if an effective vocabulary-learning method is used. Therefore, English teachers are required to teach them repeatedly (Nation, 2013). Indeed, repetition plays a crucial role in vocabulary learning. Webb (2007a) demonstrated that participants who encountered the
target words 10 times gained more productive and receptive vocabulary knowledge than those who encountered the same target words 7 times. Those who learned them 7 times acquired more productive vocabulary knowledge than those who learned them 3 times. Hence, the more often learners encounter the target words, the more vocabulary knowledge they gain.

Although a number of studies have emphasized the importance of repetition (Zeeland & Schmitt, 2013; Zhang, 2014), many researchers have argued that repetition over a short period is not helpful for long-term retention. A typical example involves massed learning or cramming, which means studying target items intensively over a short period of time before a test (Kornell, 2009). Basically, students tend to believe that massed learning is effective (Kornell, 2009), and their high scores in an immediate test after the learning session can assure them that it will be also effective for long-term retention (Rohrer & Pashler, 2007). However, it was found that massed learning is not effective for long-term retention because items learned in cramming tend to be forgotten soon afterwards (Baddeley, 2013). Ebbinghaus (1885/1964) reported that learners forgot 74% of the learned items just one day later. This means that cramming in a single study session cannot contribute greatly to long-term retention, even if it involves repeating an item several times.

In fact, a great number of studies have revealed that distributed practice, where learners distribute learning across time, is more effective with regard to long-term retention than massed learning (Kornell, 2009; Sobel, Cepeda & Kapler, 2011; Baddeley, 2013; Kapler, Weston & Wiseheart, 2015). Distributed practice usually includes several learning sessions over a certain number of days, with an interval between each session. Students are less likely to forget the items they have learned with each additional study session after a break. Another advantage of distributed practice is that it reduces the burden of learning compared to massed learning. A light workload in each session can help students to remain focused on target items. Intervals also help them to recover from fatigue, and they can refresh their memories in every session. Therefore, Baddeley (2013) argues that “it is better to distribute your learning trials across a period of time than to mass them together in a single block of learning” (p. 68).

Distributed practice is also effective for L2 vocabulary learning. We should ensure learning involves an interval between each learning session. In the first step of L2 vocabulary learning, a large proportion of learners try to establish connections between an L2 target word and its L1 equivalent. In this process, their effort to retrieve the L1 meaning from its L2 form can be very effective (Nation, 2013). Learners should be encouraged to ensure an interval between each retrieval session. Traditionally, there are two types of spaced retrieval practice: expanding retrieval practice and equally-spaced retrieval practice. In expanding retrieval practice, students relearn an item immediately after learning it initially. The space or interval between each period of learning is then increased gradually (Karpicke & Roediger, 2007). On the other hand, equally-spaced retrieval practice refers to the technique where the interval between each learning session remains the same.
Expanding retrieval practice has been considered an effective way of enhancing long-term retention. The idea originally came from Ebbinghaus (1885/1964). In his experiment, participants forgot 42% of the learned items after 20 minutes and 56% after an hour. A day, a week and a month later, they had forgotten 74%, 77%, and 79% of the learned items respectively. As Baddeley (2013) argues, “forgetting is rapid at first but gradually slows down” (p. 99). Thus, the idea that the first review session should occur as soon as possible after the initial time of learning, and that subsequent intervals should become gradually longer, has become popular among L2 language learners and teachers.

Learners using expanding retrieval practice basically have a chance to take the first review session relatively soon after the first learning session. Therefore, the possibility of successful retrieval is high here, and gradually increasing the interval between learning sessions can help learners retain the learned items (Karpicke & Roediger, 2010). A successful retrieval experience from memory produces better long-term retention than an unsuccessful experience (Karpicke & Bauernschmidt, 2011; Nakata, in press). In contrast, however, under the equally-spaced schedule, the possibility of retrieval failure is much higher during the first review session because there is usually a longer period of time between the first learning session and the first review session than in the expanding condition. This experience of retrieval failure has a negative influence on all subsequent learning sessions (Storm, Bjork & Storm, 2010). In this respect, language learners should relearn the target words immediately after learning them the first time, and then expand the intervals in order to store the words in their long-term memory (Ikemura et al., 2006).

2. Literature Review

2.1 Comparing the Effects of Expanding versus Equally-Spaced Retrieval Practice

Several studies have compared the effects of expanding and equally-spaced retrieval practice. In fact, the results tend to contradict the assumption that expanding retrieval is more effective than equally-spaced retrieval. Some studies found that equally-spaced learning was more effective than expanding retrieval learning, and others revealed that there was no difference between the two conditions.

Cull (2000) showed mixed results. His first experiment confirmed the advantage of equally-spaced learning over expanding retrieval practice. He compared participants who learned 10 cue-target word pairs (e.g. *handsel-payment*) in the expanding condition (Day 1, 2, 4 and 7), with those who learned another 10 word pairs in the equally-spaced condition (Day 1, 3, 5 and 7). On Day 10, all the participants took the final recall test, where they were asked to write down the target word for its cue word. The result revealed that the equally-spaced group showed a higher retention rate than the expanding group (84.3% versus 72.3%). However, another experiment in this study showed no significant difference between the two groups. The first experiment was replicated under the same conditions, except that the final delayed test was conducted eight days...
after the last learning session (Day 15). There was no significant difference between the expanding and equally-spaced groups (64.3% versus 74%).

Another study which used the L2 form and its L1 translation word pairs was conducted by Kang et al. (2014). Their participants learned 60 Japanese-English word pairs under the expanding (Day 1, 3, 9 and 28) and the equally-spaced (Day 1, 10, 19 and 28) conditions. Fifty-six days after the final learning session, both groups took the final delayed post-test on Day 84. No significant difference was found between the expanding and equally-spaced conditions in the immediate test conducted on Day 28 (62% versus 65%) or in the delayed post-test conducted on Day 84 (49% versus 46%).

Therefore, no studies have shown that expanding retrieval practice is superior. A possible reason is that no previous studies have conducted the first review session soon after the first study session. In order to make expanding retrieval practice effective, the first review session should be conducted the same day as the first learning session, because most forgetting occurs within a day after the study session (Ebbinghaus, 1885/1964). However, the previous studies involved a two-day interval (Kang et al., 2014), or a one-day interval (Cull, 2000; Küpper-Tetzel, Kapler & Wiseheart, 2014) between the initial learning session and the first review session.

Our previous study (Kanayama & Kasahara, 2015) therefore attempted to compensate for this limitation. We compared the effects of the expanding (Day 1, 1, 5 and 12) with the equally-spaced schedule (Day 1, 8, 15 and 22), and asked Japanese university students to remember 20 English forms and Japanese translation pairs (e.g. ligament-jinntai). We gave the expanding group their first review session just one hour after the first study session. Both groups took each test immediately after each review session. Twenty-one days after the last learning session (Days 12 & 22), both groups took a delayed post-test on Day 33 or Day 43. The results demonstrated that the expanding group had a better score than the equally-spaced group after the first review session (72.2% versus 51.3%). However, no significant difference was found between the expanding and equally-spaced condition after the second review session (79.5% versus 71.7%), the third review session (86.8% versus 79.3%), or in the delayed post-test (54.4% versus 52.6%). The results indicated that the equally-spaced group caught up with the expanding group after the second review session. Thus, this study indicated that there are no advantages to expanding retrieval practice over equally-spaced retrieval practice.

One problem in the study by Kanayama and Kasahara (2015) can be attributed to the fact that absolute spacing was not the same between the expanding and the equally-spaced groups. Absolute spacing means the total number of intervals in the whole learning period (Nakata, in press). For example, “if a given item is encountered four times, and each encounter is separated by 2 min, the absolute spacing is 6 min (2 min×3)”. Nakata argues that longer absolute spacing can be more effective in learning. For instance, Karpicke and Bauernschmidt (2011) compared different absolute spacing conditions; short spacing (1-5-9; 5-5-5), medium spacing (5-10-15; 10-10-10) and long spacing (15-30-45; 30-30-30). No significant difference was found between
the expanding and equally-spaced schedules, but a significant difference was found between absolute spacing schedules. Long spacing produced a higher retention rate than medium spacing (75% versus 64%), and in turn, medium spacing had better results than short spacing (64% versus 49%). Increasing the absolute spacing of repeated retrievals improves long-term retention. In the case of Kanayama and Kasahara (2015), there was longer absolute spacing in the equally-spaced group (21 days) than in the expanding group (11 days). This longer absolute spacing in the equally-spaced group may have been of considerable benefit to the participants. Therefore, the purpose of the present study was to replicate our previous study, with the same absolute spacing given to each condition.

2.2 Hypotheses

Considering the results of previous studies (Kang et al., 2014; Kanayama & Kasahara, 2015), the present study assumed that expanding retrieval practice would be more beneficial than equally-spaced retrieval practice after the first review session. However, equally-spaced retrieval practice should catch up with expanding retrieval practice as long as the learners in the equally-spaced group have more than one review session. This study set up the following four hypotheses:

(1) In the immediate recall test after the first review session, the expanding schedule shows a better score than the equally-spaced schedule.

(2) In the immediate recall test after the second review session, there is no difference in score between the expanding schedule and the equally-spaced schedule.

(3) In the immediate recall test after the third review session, there is no difference in score between the expanding schedule and the equally-spaced schedule.

(4) In the final delayed recall test, there is no difference in vocabulary retention between the expanding schedule and the equally-spaced schedule.

3. Method

3.1 Participants

The participants were 83 first-grade university students who belonged to two English classes. All the learners had studied English for a minimum of six years. However, 30 participants were removed from the data analysis because they were absent from at least one of the experimental sessions. Hence, data for 53 participants were surveyed. The expanding retrieval group (Day 1, 1, 8 and 22) was designated Group A (n = 34) and the equally-spaced retrieval group (Day 1, 8, 15 and 22) Group B (n = 19).

In order to investigate whether there was any difference in vocabulary size between the groups, the Vocabulary Size Test (Nation & Beglar, 2007) was carried out for both groups at the 1000, 2000 and 3000 word levels. The mean score of Group A ($M = 19.91, SD = 2.41$) was quite
similar to that of Group B ($M = 19.84$, $SD = 3.07$). An independent $t$ test found no significant difference between Group A and Group B, $t (51) = .91$, $p = .928$, $r = .012$. This confirmed that there was no significant difference in vocabulary size between the groups.

3.2 Materials

The present study used L2 low-frequency English words and L1 Japanese equivalent pairs for an educational reason. In order to use English words that were not known to the participants, 50 low-frequency words consisting of 25 nouns and 25 verbs were presented to them before the experiment sessions as a pre-test. The participants had to say whether they knew any of the 50 words (see Appendix B). We selected 20 words, which none of the participants knew, as the target words in the experiment. These target words comprised 10 nouns and 10 verbs which were not listed in JACET 8000 (JACET, 2003). All 10 nouns and verbs are shown in Table 1 with their Japanese translations.

Table 1

<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>(ネコが)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>靭帯</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>忠告する</td>
</tr>
</tbody>
</table>

3.3 Procedures

First, both groups were given the Vocabulary Size Test to measure their vocabulary size and to confirm that there was no difference in vocabulary size between the groups. Second, a week after the Vocabulary Size Test, they were given a word list containing the 20 English target words with their Japanese translations. A blank sheet of paper was also given out in case the participants wished to memorize the target words by writing them. They were then asked to remember as many target items as possible in five minutes after receiving instructions on how to pronounce each item. Afterwards, the first immediate recall test session took place, where they were required to write down the meaning of each word in Japanese within five minutes. The
participants did the same thing for each subsequent test. The order of the items was randomized on each testing occasion.

For the participants in Group A, the second learning session (= the first review session) and the second immediate test occurred only an hour after the first learning session. The third learning session (= the second review session) and third immediate test were run seven days after the second learning session. Fourteen days after the third learning session, the participants attended the fourth learning session (= the third review session) and fourth immediate test. Finally, the delayed post-test was carried out without the five-minute learning session and with no advance notice, 21 days after the fourth learning session.

For the participants in Group B, however, the interval between sessions remained the same. Specifically, the second learning session (= the first review session) and the second immediate test were conducted a week after the first learning session, and the third learning session (= the second review session) and third immediate test a week after the second learning session. In the same way, the fourth learning session (= the third review session) and immediate test were carried out a week after the third learning session. Twenty-one days after the fourth learning session, the participants took the delayed post-test with no learning session beforehand. Table 2 presents the schedule for each learning session and immediate test, and for the delayed post-test.

### Table 2

*Schedules for learning sessions and immediate tests, and for the post-test*

<table>
<thead>
<tr>
<th>Group A (Expanding)</th>
<th>Group B (Equally-spaced)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>Day 1</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; learning session and 1&lt;sup&gt;st&lt;/sup&gt; test</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; learning session and 1&lt;sup&gt;st&lt;/sup&gt; test</td>
</tr>
<tr>
<td>An hour interval</td>
<td>7-day interval</td>
</tr>
<tr>
<td>Day 1</td>
<td>Day 8</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; learning session and 2&lt;sup&gt;nd&lt;/sup&gt; test</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; learning session and 2&lt;sup&gt;nd&lt;/sup&gt; test</td>
</tr>
<tr>
<td>7-day interval</td>
<td>7-day interval</td>
</tr>
<tr>
<td>Day 8</td>
<td>Day 15</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; learning session and 3&lt;sup&gt;rd&lt;/sup&gt; test</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; learning session and 3&lt;sup&gt;rd&lt;/sup&gt; test</td>
</tr>
<tr>
<td>14-day interval</td>
<td>7-day interval</td>
</tr>
<tr>
<td>Day 22</td>
<td>Day 22</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; learning session and 4&lt;sup&gt;th&lt;/sup&gt; test</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; learning session and 4&lt;sup&gt;th&lt;/sup&gt; test</td>
</tr>
<tr>
<td>21-day interval</td>
<td>21-day interval</td>
</tr>
<tr>
<td>Day 43</td>
<td>Day 43</td>
</tr>
<tr>
<td>Delayed post-test</td>
<td>Delayed post-test</td>
</tr>
</tbody>
</table>

### 3.4 Scoring

Two points were given for each correct answer and one point was given for a partially correct answer. Where the participants made a mistake in terms of a part of speech, this was considered a partially correct answer. For example, in the case of the word “encroachment”, the correct answer was *shinryaku*, and one point was given for *shinryaku-suru*. Another example of a partially correct answer was where an intransitive verb was given for a transitive verb (e.g. *okoru*
for “exasperate”) and vice versa. The maximum score for each test was 40 points (2 points × 20 words).

3.5 Data Analysis

A 2 (Group: Group A, Group B) × 5 (Test: Immediate Test 1, 2, 3, 4, Delayed Test 1) mixed ANOVA was conducted in order to examine Hypotheses 1 to 4. In addition, a simple main effect test was carried out between the mean scores of Group A and Group B in each immediate and delayed test respectively. Post-hoc multiple comparisons were also carried out between all tests in Group A and between all tests in Group B.

4. Results

Table 3 presents the mean scores and standard deviations of each immediate test (Immediate Test 1, 2, 3, 4) and the delayed post-test (Delayed Test 1) for Group A and Group B. Figure 1 shows the mean scores in each immediate and delayed test for Group A and Group B. A 2 (Group: Group A, Group B) × 5 (Test: Immediate Test 1, 2, 3, 4, Delayed Test 1) mixed ANOVA showed that there was significant interaction between the two factors, Group and Test, $F(1, 51) = 4.619$, $p = .001$, $\eta^2 = .083$, and that there was a main effect of Test, $F(1, 51) = 129.85$, $p < .001$, $\eta^2 = .718$, but no main effect of Group, $F(1, 51) = 2.37$, $p = .13$, $\eta^2 = .044$.

Furthermore, there was a simple main effect between the two groups in Immediate Test 2, $F(1, 51) = 10.861$, $p < .005$, but no simple main effect between the two groups in Immediate Test 1, $F(1, 51) = 1.66$, $p = .198$, Immediate Test 3, $F(1, 51) = .503$, $p = .479$, Immediate Test 4, $F(1, 51) = .342$, $p = .559$ or Delayed Test 1, $F(1, 51) = .891$, $p = .346$ respectively. This means that the expanding group had a better score than the equally-spaced group in Immediate Test 2 (70.3% versus 51.3%), that the expanding group showed the same effects as the equally-spaced group in Immediate Test 1 (40.8% versus 33.4%), Immediate Test 3 (75.8% versus 71.7%), Immediate Test 4 (82.7% versus 79.3%), and that the expanding group showed the same retention rates as the equally-spaced group in Delayed Test 1 (58.1% versus 52.6%).

Post-hoc multiple comparisons were carried out between all the tests in Group A. Significant differences were found between Immediate Test 1 and Immediate Test 2, $t(33) = 11.63$, $p = .00$, between Immediate Test 2 and Immediate Test 3, $t(33) = 2.14$, $p = .032$, and between Immediate Test 3 and Immediate Test 4, $t(33) = 2.72$, $p = .006$. This means that the more learning and test sessions the participants had, the better their performance.

Furthermore, post-hoc multiple comparisons were administered between all the tests in Group B. Significant differences were found between Immediate Test 1 and Immediate Test 2, $t(18) = 5.28$, $p = .00$, between Immediate Test 2 and Immediate Test 3, $t(18) = 6.01$, $p = .00$, and between Immediate Test 3 and Immediate Test 4, $t(18) = 2.25$, $p = .025$. As with Group A, the more learning and test sessions the participants in Group B experienced, the higher their scores.
Table 3
Means and standard deviations of each immediate and delayed test for Group A and Group B (N = 53, Full score = 40)

<table>
<thead>
<tr>
<th></th>
<th>Group A (expanding, n = 34)</th>
<th>Group B (equally-spaced, n = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Immediate Test 1</td>
<td>16.35</td>
<td>7.76</td>
</tr>
<tr>
<td>Immediate Test 2</td>
<td>28.15</td>
<td>8.03</td>
</tr>
<tr>
<td>Immediate Test 3</td>
<td>30.32</td>
<td>8.11</td>
</tr>
<tr>
<td>Immediate Test 4</td>
<td>33.09</td>
<td>7.48</td>
</tr>
<tr>
<td>Delayed Test 1</td>
<td>23.24</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Figure 1. The means of each immediate and delayed test for Group A and Group B

5. Discussion

Hypothesis 1 posits that the expanding schedule shows a better score than the equally-spaced schedule after the first review session. This was supported because there was a significant difference between the two groups in Immediate Test 2 (70.3% versus 51.3%). This result was consistent with Kanayama and Kasahara (2015), and confirmed a conventional idea that learners benefit significantly from the first review session soon after the first learning session.

Hypothesis 2 states that, in the immediate recall test, there is no difference in score between the expanding schedule and the equally-spaced schedule after the second review session. This was also supported because no significant difference was observed between the two groups in Immediate Test 3 (75.8% versus 71.7%). It is conceivable that, since Group A had already achieved a relatively high score in Immediate Test 2 (70.3%), there may have been little room for improvement. On the other hand, the equally-spaced group had a lower score than the expanding
group in Immediate Test 2 (51.3%). Thus, there could be sufficient room for Group B to improve their performance. Indeed, the equally-spaced group improved their performance by 20.4% (71.7%—51.3%). However, the score of the expanding group increased by 5.5% (75.8%—70.3%).

Hypothesis 3 assumes that, in the immediate recall test, there is no difference in score between the expanding schedule and the equally-spaced schedule after the third review session. This was supported, since no significant difference was found between the two groups in Immediate Test 4 (82.7% versus 79.3%), corresponding with the previous findings (Kang et al., 2014; Kanayama & Kasahara, 2015). Both groups showed the same performance in Immediate Test 3, which led to the same recall rate in Immediate Test 4. Consequently, it was indicated that the scores of the equally-spaced group had caught up with those of the expanding group since Immediate Test 3. In fact, little difference was found between the two groups in that sufficient repetition and total learning time were ensured in both.

Hypothesis 4 predicts that the expanding schedule shows the same vocabulary retention rates as the equally-spaced schedule in the delayed post-test. This was confirmed because there was no significant difference between the two groups in Delayed Test 1 (58.1% versus 52.6%). A possible reason for this result was the use of the same interval after the last learning sessions, 21 days. Both groups showed the same performance rate in Immediate Test 4, which led to the same retention rate in Delayed Test 1. Therefore, this study replicated the previous studies (Kang et al., 2014, Kanayama & Kasahara, 2015), and revealed that expanding retrieval practice was as effective as equally-spaced retrieval practice in long-term retention of L2 vocabulary, even when the same absolute spacing was employed in each condition.

The findings of this study indicate that a decisive factor in enhancing long-term retention of L2 vocabulary is the number of repetitions, not the length of spacing. Repetitions can consolidate connections between L2 target words and their L1 equivalents. It is true that the first review session should be done soon after the first learning session because learners forget most within a day after the learning session. However, the equally-spaced group in this study, who had the first review session seven days after the first learning session, caught up with the score of the expanding group after the second review session. This means that the second review session compensated for the disadvantage of the delayed first review session. Indeed, the equally-spaced group benefited more from the second review session than the expanding group (20.4% versus 5.5%). Once the equally-spaced group had caught up with the expanding group, no subsequent difference was observed. It follows that if learners are given several learning sessions, the spaces between the sessions do not significantly affect their total benefit from the learning. Hence, repetitions are a crucial factor in enhancing long-term retention of L2 and L1 connections.

How should English teachers apply the results of the present and previous studies to L2 vocabulary instructions? There are two pedagogic implications. First, English teachers should carry out the first review session soon after the first study session. For instance, giving learners
five minutes’ vocabulary instruction at the beginning and a review session at the end of an English lesson would be more effective than giving them five minutes of vocabulary instruction in one lesson and reviewing it one day later. The learners would spend much less time and effort relearning the target items. By giving the first review soon after the first learning session, the teachers should be able to ensure their students do not forget the words they have learned.

Second, teachers should consider giving their students repeated instances of the same words throughout a given time period, rather than considering how each learning session should be spaced. Indeed, learners in both the expanding and equally-spaced condition benefit greatly from repetition. However, equally-spaced retrieval practice seems to be more applicable to the actual classroom situation because a weekly lesson schedule is usually fixed in many schools. With a rigid class schedule in the school setting, it would be difficult for English teachers to fit an appropriate expanding interval timescale into L2 vocabulary teaching, and to make use of the expanding retrieval method. The equally-spaced retrieval method would, however, be easy to implement because each English lesson is conducted between regular intervals.

Finally, there are some limitations to be noted in the present study. First, it involved a small number of participants. In particular, Group B had only 19 participants. Further study is required to collect a greater volume of participant data. Second, only 20 English and Japanese pairs were used in this study. In the last stage of the learning session, a ceiling effect was observed in both Group A and Group B in Immediate Test 4 (82.7% versus 79.3%). Further studies should require participants to remember more target words, and this may highlight potential differences between the expanding and the equally-spaced groups. Third, we gave our participants a receptive vocabulary test (L2→L1), but not a productive vocabulary test (L1→L2). Hence, we should investigate the effects of L2 form retrieval in a further study.

Fourth, there is some doubt whether the expanding condition was conducted in a truly gradual way. The four learning sessions were arranged in the following schedule: Day 1, an hour later, Day 8 and Day 22. The spacing between each learning session became longer quite rapidly. However, expanding retrieval practice involves a gradual interval increase between each learning experience (Karpicke & Roediger, 2010). Thus, further studies should increase each interval more gradually in the expanding group.

6. Conclusions

One of the purposes of this study was to challenge the assumption by Ebbinghaus that students should learn items at expanding intervals. The finding of the present study is of some value because this seems to be the first study in L2 vocabulary acquisition research to show no significant difference between expanding and equally-spaced retrieval practice. The main purpose of this study was to compare the effects of expanding and equally-spaced retrieval practice on L2 vocabulary learning. The most crucial factor contributing to strong L2 and L1 connections is the
number of repetitions, not the length of the interval between each learning session. As long as learners are given the chance to have four learning sessions, both learning schedules ultimately show the same performance.

References


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


Appendix A: Learning Sheet (For Group A and Group B)

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: 鞄帯(じんたい)。

hoof: (馬などの)ひづめ。

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 and Group B)

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