

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

Tatsuya Nakata

Faculty of Foreign Language Studies, Kansai University

doi: <http://dx.doi.org/10.7820/vli.v04.1.nakata>

Abstract

Although the effects of retrieval-based learning may be affected by factors such as relative spacing, absolute spacing, and feedback timing, few studies have examined learners' perceived effects of these factors on second language (L2) vocabulary learning. With this limitation in mind, the present study examined learners' perceived effects of the above three factors on L2 vocabulary learning. A questionnaire was administered to 226 Japanese college students. The results showed that the participants (1) considered relative spacing not to affect learning, (2) perceived spaced learning to be more effective than massed learning, and (3) considered immediate and delayed feedback to be equally effective, all of which are supported by earlier empirical research. The results suggest that L2 learners are able to make accurate judgments about the effectiveness of retrieval-based vocabulary learning techniques.

1 Introduction

Research shows that retrieval enhances second language (L2) lexical acquisition (e.g., Barcroft, 2007; Karpicke & Roediger, 2008). Retrieval refers to recalling information about L2 words from memory. Earlier studies also suggest that the effects of retrieval-based learning may be affected by factors such as relative spacing, absolute spacing, and feedback timing (see below for definitions). Relative spacing is defined as how retrieval attempts are distributed relative to one another (Karpicke & Bauernschmidt, 2011). Existing studies have examined the effects of two types of relative spacing schedules: equal and expanding. In the former, spacing between retrievals of a given word is held constant (e.g., 4, 4, and 4 days). In the latter, spacing between retrievals is gradually expanded (e.g., 1, 4, and 7 days). Some researchers argue that expanding spacing is more effective than equal spacing (e.g., Baddeley, 1997; Ellis, 1995; Hulstijn, 2001; Nation, 2001; Schmitt, 2007; see Roediger & Karpicke, 2010, for a review). Although some non-L2 vocabulary studies have found the advantage of expanding over equal spacing (e.g., Cull, Shaughnessy, & Zechmeister, 1996; Landauer & Bjork, 1978; Storm, Bjork, & Storm, 2010), L2 vocabulary studies have found little difference between the effects

of the two types of relative spacing schedules (Kang, Lindsey, Mozer, & Pashler, 2014; Karpicke & Bauernschmidt, 2011; Nakata, in press; Pyc & Rawson, 2007).

Not only relative but also absolute spacing is found to affect vocabulary learning. Absolute spacing is defined as the total amount of spacing that intervenes all retrieval opportunities of a given word (Karpicke & Bauernschmidt, 2011). For example, if a given word is retrieved five times, and each retrieval is intervened by 4 minutes, absolute spacing is 16 minutes (4 minutes \times 4). There are two phenomena regarding the effects of absolute spacing: spacing effect and lag effect. First, according to the spacing effect, spaced learning, which introduces spacing between retrievals of a given word, increases learning more than a massed schedule, which does not involve any spacing (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1989, 1996; Janiszewski, Noel, & Sawyer, 2003, for meta-analyses). Second, the lag effect refers to a phenomenon where larger absolute spacing enhances learning more than shorter absolute spacing (e.g., Bird, 2010; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Cepeda et al., 2006, 2009; Pashler, Rohrer, Cepeda, & Carpenter, 2007; Rohrer & Pashler, 2007). Although the spacing effect has been found to affect L2 vocabulary acquisition (e.g., Karpicke & Bauernschmidt, 2011; Nakata, in press), existing L2 vocabulary studies have produced inconsistent results regarding the lag effect (e.g., Crothers & Suppes, 1967; Karpicke & Bauernschmidt, 2011; Nakata, in press; Nakata & Webb, in press; Pashler, Zarow, & Triplett, 2003; Pyc & Rawson, 2007).

Another line of research suggests that the timing of feedback after retrieval attempts affects learning (e.g., Butler, Karpicke, & Roediger, 2007; Kulik & Kulik, 1988; Metcalfe, Kornell, & Finn, 2009; Nakata, 2015). For example, suppose that the learner was asked to retrieve the meaning of a Japanese word *inu* (dog). Which would be more effective, to provide the correct answer 3 seconds after the retrieval attempt (immediate feedback) or 1 hour after (delayed feedback)? Some researchers claim that delayed feedback increases learning more than immediate feedback because the former introduces larger spacing than the latter, which should facilitate learning according to the lag effect (e.g., Butler et al., 2007; Metcalfe et al., 2009). Although some non-L2 vocabulary studies have found the facilitative effects of delaying feedback (e.g., Butler et al., 2007; Kulik & Kulik, 1988; Metcalfe et al., 2009), L2 vocabulary research has failed to support this finding (Nakata, 2015).

The findings of earlier research are useful because they allow us to determine the optimal way to learn L2 vocabulary from retrieval. One limitation of existing research, however, is that none of them examined learners' perceived effects of relative spacing, absolute spacing, and feedback timing on L2 vocabulary learning. Cognitive psychology literature on judgments of learning (JOLs) suggests that learners are often unable to make accurate judgments about the effectiveness of retrieval-based learning techniques (e.g., Karpicke & Roediger, 2008; Kornell, 2009; Kornell & Son, 2009; Roediger & Karpicke, 2006). Examining learners' perceived effectiveness of retrieval-based learning techniques is of value because learners' perceptions may encourage or discourage the use of effective learning techniques. For instance, learners perhaps tend to employ vocabulary learning techniques that they consider more effective than those they consider less effective, as long as other variables (e.g., time or cost

required) are held constant. With the above discussion in mind, the present study examined learners' perceived effects of relative spacing, absolute spacing, and feedback timing on L2 vocabulary learning.

2 Method

A total of 226 Japanese college students participated in this study and studied English-Japanese word pairs using flashcard software (e.g., Nakata, 2011, 2013) developed by the author. One hundred and thirty-eight of them were randomly assigned to one of the four groups: massed, short, medium, and long spacing. In all four groups, the treatment involved typing the English target word corresponding to a Japanese translation provided (e.g., 幽霊 = ____?). In the massed group, 20 target English-Japanese word pairs (e.g., *apparition* – 幽霊) were practiced four times in a row without any spacing. In the short, medium, and long spacing groups, four encounters of a given target word pair were separated by approximately 1, 2, and 6 minutes on average, respectively. In the short, medium, and long spacing groups, half of the target word pairs were studied in an expanding spacing condition, and the other half were studied in an equal spacing condition. There was no equal or expanding spacing condition in the massed group because all target items were encountered four times in a row, and there was no spacing in this group. After completing the treatment and posttests, a questionnaire was administered to examine learners' perceived effects of relative and absolute spacing schedules on learning. The short, medium, and long spacing groups were asked to evaluate the usefulness of the equal and expanding schedules for learning on a 7-point scale, where 1 means *Not helpful at all* and 7 means *Very helpful*. Participants in the massed group were asked to evaluate the usefulness of the massed schedule for learning on the same 7-point scale.

The remainder of the participants studied 16 English-Japanese word pairs using flashcard software. The treatment for these participants also involved typing the English target word corresponding to a Japanese translation provided (e.g., 幽霊 = ____?). For a given participant, for half of the target word pairs, the target English-Japanese word pair (e.g., 幽霊 = *apparition*) was provided as feedback immediately after each retrieval opportunity (immediate feedback). For the other half, feedback was not given until all other items were studied (delayed feedback). After completing the treatment and posttests, the participants were asked to evaluate the effectiveness of immediate and delayed feedback on a 7-point scale, where 1 means *I learned more with immediate feedback* and 7 means *I learned more with delayed feedback*.

3 Results

Regarding the perceived effects of relative spacing schedules, the average rating (*SDs* in parentheses) on a 7-point scale (1: *Not helpful at all*; 7: *Very helpful*) was 5.16 (1.08), 5.09 (1.30), 4.91 (1.78), 4.75 (1.88), 4.55 (1.59), and 4.65 (1.40) for the short equal, short expanding, medium equal, medium expanding, long equal, and long expanding conditions, respectively. (One participant in the long spacing group did not provide responses.) When collapsed across the absolute spacing

groups, the average rating was 4.87 (1.52) for equal spacing and 4.83 (1.55) for expanding spacing. The responses were entered into a two-way mixed design 2 (relative spacing: equal/expanding) \times 3 (absolute spacing: short/medium/long) ANOVA. None of the main effects or interaction was significant, main effect of relative spacing: $F(1, 92) = 0.14, p = .706, \eta_p^2 < .01$; main effect of absolute spacing: $F(2, 92) = 1.07, p = .349, \eta_p^2 = .02$; interaction between the two variables: $F(2, 92) = 0.47, p = .627, \eta_p^2 = .01$. The results are also supported by the small effect sizes ($\eta_p^2 \leq .02$). The findings indicate that (1) no statistically significant difference existed between equal and expanding spacing in the ratings, and (2) the three groups did not differ significantly from each other in their responses regarding the effects of equal and expanding spacing.

Next, let us examine learners' perceived effects of the four absolute spacing schedules (massed, short, medium, and long) on learning. In the short, medium, and long groups, the scores for equal and expanding spacing were averaged out and compared with those for the massed schedule. For instance, if a participant in the short spacing group gave 5 for equal and 6 for expanding spacing, 5.5 was used as this participant's rating for the short spacing schedule. Because there was little difference between the ratings for equal ($M = 4.87, SD = 1.52$) and expanding spacing ($M = 4.83, SD = 1.55$; see above), it was judged appropriate to use the mean score of the two relative spacing schedules to represent a given absolute spacing schedule. The average rating (SD s in parentheses) on a 7-point scale (1: *Not helpful at all*; 7: *Very helpful*) was 3.78 (1.41), 5.13 (1.09), 4.83 (1.73), and 4.60 (1.44) for the massed, short, medium, and long spacing schedules, respectively. A one-way ANOVA found a statistically significant difference among the four groups, $F(3, 123) = 5.18, p = .002, \eta^2 = .01$. According to the Bonferroni method of multiple comparisons, the scores for the short and medium spacing schedules were significantly higher than those for the massed schedule, producing medium to large effect sizes (massed vs. short: $p = .002, d = 1.07$; massed vs. medium: $p = .025, d = 0.66$). The difference between the massed and long schedules did not reach statistical significance ($p = .154$), but a medium effect size was observed ($d = 0.59$). The differences were not statistically significant for all other comparisons ($p \leq 1.000$), and no more than small effect sizes were observed ($0.15 \leq d \leq 0.42$). These results suggest that (1) the participants perceived the short and medium spacing schedules to be more effective than the massed schedule, and (2) the short, medium, and long spacing groups did not differ significantly from each other in their responses.

As for the effects of feedback timing, the average rating on a 7-point scale (1: *I learned more with immediate feedback*; 7: *I learned more with delayed feedback*) was 3.85 ($SD = 2.01$). The results indicate that learners tended to believe that they learned equally well from the two types of feedback.

4 Discussion and Conclusions

The present study showed that the participants (1) considered equal and expanding spacing to be equally effective, (2) perceived spaced learning to be more effective than massed learning, and (3) considered immediate and delayed feedback to be equally effective, all of which are supported by existing empirical research

(e.g., Kang et al., 2014; Karpicke & Bauernschmidt, 2011; Nakata, 2015, in press; Pyc & Rawson, 2007). The results suggest that contrary to the findings of earlier research on judgments of learning (e.g., Karpicke & Roediger, 2008; Kornell, 2009; Kornell & Son, 2009; Roediger & Karpicke, 2006), the participants in this study were able to make accurate judgments about the effectiveness of retrieval-based learning techniques. The inconsistent results were perhaps due to a difference in participants. Although most existing studies on judgments of learning have been conducted with American university students (e.g., Karpicke & Roediger, 2008; Kornell, 2009; Kornell & Son, 2009; Roediger & Karpicke, 2006), this study was conducted with Japanese college students. Because Japanese students tend to be more experienced in and proficient at rote memorisation than American students (Tinkham, 1989), the participants of the present research were perhaps more familiar with effective learning strategies than suggested by earlier studies. This may be partly the reason why the results of this study were at odds with those of the existing literature on judgments of learning. However, due to the paucity of metacognition research on Japanese students, further research is warranted.

Although the findings of this study are useful, the present study also suffers from some limitations. One limitation is that the questionnaire was given only to Japanese college students. Future research may replicate this study with participants from different backgrounds. Another limitation is that the current study only examined learners' perceived effectiveness of retrieval-based learning techniques after the treatment. In future research, it would be useful to give a questionnaire before and after the treatment.

References

- Baddeley, A. D. (1997). *Human memory: Theory and practice* (Revised ed.). East Sussex: Psychology Press.
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57, 35–56. doi:10.1111/j.1467-9922.2007.00398.x
- Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, 31, 635–650. doi:10.1017/S0142716410000172
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273–281. doi:10.1037/1076-898X.13.4.273
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, 56, 236–246. doi:10.1027/1618-3169.56.4.236
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380. doi:10.1037/0033-2909.132.3.354

- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridge of optimal retention. *Psychological Science, 19*, 1095–1102. doi:10.1111/j.1467-9280.2008.02209.x
- Crothers, E., & Suppes, P. (1967). *Experiments in second language learning*. New York, NY: Academic Press.
- Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2*, 365–378. doi:10.1037/1076-898X.2.4.365
- Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*, 309–330. doi:10.1007/BF01320097
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Human memory* (pp. 197–236). San Diego, CA: Academic Press.
- Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*, 103–128. doi:10.1080/0958822940080202
- 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: Cambridge University Press.
- Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*, 138–149. doi:10.1086/374692
- 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. doi: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. doi:10.1037/a0023436
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968. doi:10.1126/science.1152408
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*, 1297–1317. doi:10.1002/acp.1537
- Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*, 493–501. doi:10.1080/09658210902832915
- Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79–97. doi:10.3102/00346543058001079

- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
- Metcalf, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, *37*, 1077–1087. doi:10.3758/MC.37.8.1077
- Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*, 17–38. doi:10.1080/09588221.2010.520675
- Nakata, T. (2013). Web-based lexical resources. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 6166–6177). Oxford: Wiley-Blackwell.
- Nakata, T. (2015). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*, *19*, 416–434. doi:10.1177/1362168814541721
- Nakata, T. (in press). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*. doi:10.1017/S0272263114000825
- Nakata, T., & Webb, S. A. (in press). 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*. doi:10.1017/S0272263115000236
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin and Review*, *14*, 187–193. doi:10.3758/BF03194050
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1051–1057. doi:10.1037/0278-7393.29.6.1051
- Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927. doi:10.3758/BF03192925
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x
- Roediger, H. L., & Karpicke, J. D. (2010). Intricacies of spaced retrieval: A resolution. In A. S. Benjamin. (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 23–47). New York, NY: Psychology Press.

- Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, 16, 183–186. doi:10.1111/j.1467-8721.2007.00500.x
- Schmitt, N. (2007). Current trends in vocabulary learning and teaching. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (pp. 827–842). Norwell, MA: Springer.
- Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, 38, 244–253. doi:10.3758/MC.38.2.244
- Tinkham, T. (1989). Rote learning, attitudes, and abilities: A comparison of Japanese and American students. *TESOL Quarterly*, 23, 695–698. doi:10.2307/3587547