Measuring Attrition of L2 Receptive Vocabulary Knowledge Over the Summer Vacation

Brandon Kramer  Brandon.L.Kramer@gmail.com
Tohru Matsuo  tohruwombat@me.com
Stuart McLean  stumc93@gmail.com
Steve Cornwell  cornwell@wilmina.ac.jp

Background

Most students at Japanese universities attend classes for about eight months of the year, with breaks during the summer and early spring for around two months each. During this time students often take a break from studying and focus on working part-time jobs, relaxing with friends, or travelling. When removed from the school context rich in foreign language stimulus, students might be expected to lose access to some of the knowledge they have gained, in a process called attrition. The degree of lexical attrition during these breaks is of great concern, and any interventions found to impede this attrition would be valuable for all stakeholders.

Aims

While previous researchers have studied summer L2 lexical attrition (for summaries see Bardovi-Harlig & Stringer, 2010; Grendel, Weltens, & de Bot, 1993), very few have examined the effects of targeted interventions on preventing this loss. This study aims to address that gap by investigating the state of L2 receptive vocabulary knowledge before and after summer vacation. Specifically, we were interested in the following treatments: (a) digital paired-associate studying of the target words, which would promote access through spaced retrieval (Nakata, 2011); (b) extensive reading, because “literacy supports retention and impedes attrition” (Bardovi-Harlig & Stringer, 2010, p. 25); and (c) travelling abroad, which should increase exposure to the L2 target vocabulary.

Research Questions:

RQ1: Can attrition be found in Japanese university students’ L2 receptive vocabulary knowledge after a two-month summer break?
RQ2: Is there a relationship between changes in L2 receptive vocabulary knowledge and time spent using a digital paired-associate vocabulary study application, words read through extensive reading, or experience travelling abroad?
Methods

Context

The context for this study was a private women’s university in western Japan with an institution-wide vocabulary program that encourages the deliberate studying of the most frequent words of English from the New General Service List (NGSL; Browne, Culligan, & Phillips, 2013). Using a freely available web application (Memrise.com), students studied the NGSL words in a paired-associate format utilizing spaced repetition (for more details on the benefits of such applications, see Nakata, 2011). Throughout the first semester, the students studied a continually expanding range of words (60 new words each week, for 13 weeks) and were tested with two productive review quizzes each week (one written and one spoken) within different classes of an integrated curriculum. At the end of the first semester, all first-year students had reviewed, in principle, the first 1,280 words of the NGSL (not including the first 500 words, which were only studied by the remedial classes). In addition to the vocabulary program, all first-year students were required to complete an extensive reading requirement (180,000 words were necessary for full points) using the web application Xreading.com, where they were exposed to similar words to those studied from the NGSL. After this first semester finished, the students were encouraged to review studied words with the Memrise app and read books from Xreading.com in their free time over the summer vacation.

Procedures

Three target vocabulary words were chosen randomly from each of the weekly ranges. Two rounds of tests measuring these target words ($K = 39$) were given to four intact classes of first-year Japanese university students, and only those participants who were present for both administrations were retained for analysis ($N = 103$; 18-19 years old; TOEIC range: 155-550). For each administration, a meaning-recall test was given first, followed by a four-choice meaning-recognition (i.e., multiple-choice) format (see example items in Figures 1 & 2). This order was chosen to avoid a test-effect where the participants could learn the meanings from the multiple-choice distractors, influencing the results of the meaning-recall test. The tests were administered digitally via SurveyMonkey.com, and the questions within each test format were randomized for each participant to reduce cheating. Participants were told they could skip questions they did not know the answer to, and each administration took about 30 minutes. The first administration was given between July 23rd and July 27th, 2018, and the second administration was given immediately upon the start of the second semester, September 24th to 28th.
Time spent studying vocabulary during the final month of summer vacation was found for each student through Memrise. Unfortunately, more detailed information than this was unavailable, because the application lacked a dedicated Learner Management System. Words read through the school’s extensive reading program were measured using the word counts from Xreading. As per the system’s settings, only the words for those books the student was able to demonstrate adequate understanding of were counted, meaning they were able to score at least 60% on the five-item comprehension check. Finally, travelling abroad information was self-reported using a questionnaire given following the final test administration in September. The students were asked to list any trips they took to a foreign country, and any which were English-speaking countries were dummy-coded for analysis.

Results

The meaning-recall data were scored dichotomously by a single native speaker of Japanese, also fluent in English. The items were scored according to whether the participants could demonstrate knowledge of the target vocabulary as used in the example sentences. For instance, even if the core meaning of the word was correct, it was not awarded a point when the part of speech was incorrect. An additional rater rated 10 randomly selected questions to conduct a rater reliability analysis ($\kappa = .618$; agreement = 81.0%). Descriptive statistics for all tests can be seen in Table 1.

For both the recall and recognition tests, a paired-samples t-test was conducted using JASP (JASP Team, 2018) to determine if there was a difference between pre- and post-test measures (see Table 2 and Figure 3). A significant difference ($p < .05$) was found between the meaning recognition tests, with the participants scoring slightly higher on the post-test ($M = 32.5$) than the pre-test ($M = 31.7$), indicating a gain in knowledge.
Table 1

Pre- and Post-Test Scores for Meaning Recall and Meaning Recognition Vocabulary Tests

<table>
<thead>
<tr>
<th>Vocabulary Measure</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SEM</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall Pre-Test</td>
<td>4</td>
<td>35</td>
<td>18.8</td>
<td>0.73</td>
<td>7.36</td>
<td>0.10</td>
<td>-0.67</td>
<td>.877</td>
</tr>
<tr>
<td>Recall Post-Test</td>
<td>4</td>
<td>37</td>
<td>18.4</td>
<td>0.72</td>
<td>7.27</td>
<td>0.22</td>
<td>-0.52</td>
<td>.872</td>
</tr>
<tr>
<td>Recognition Pre-Test</td>
<td>20</td>
<td>39</td>
<td>31.7</td>
<td>0.46</td>
<td>4.72</td>
<td>-0.60</td>
<td>-0.33</td>
<td>.811</td>
</tr>
<tr>
<td>Recognition Post-Test</td>
<td>19</td>
<td>39</td>
<td>32.5</td>
<td>0.44</td>
<td>4.52</td>
<td>-0.99</td>
<td>0.63</td>
<td>.816</td>
</tr>
</tbody>
</table>

Note. N = 103; SES (standard error of skewness) = 0.24 for all tests; SEK (standard error of kurtosis) = 0.47 for all tests; α = Cronbach’s alpha.

Table 2

Paired Samples T-Tests between Pre- and Post- Measures

<table>
<thead>
<tr>
<th>Vocabulary Measure</th>
<th>t</th>
<th>p</th>
<th>M</th>
<th>SE</th>
<th>Low</th>
<th>High</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall Pre-Post</td>
<td>0.94</td>
<td>0.350</td>
<td>0.38</td>
<td>0.403</td>
<td>-0.42</td>
<td>1.18</td>
<td>0.09</td>
</tr>
<tr>
<td>Recognition Pre-Post</td>
<td>-3.10</td>
<td>0.002</td>
<td>-0.86</td>
<td>0.279</td>
<td>-1.42</td>
<td>-0.31</td>
<td>-0.31</td>
</tr>
</tbody>
</table>

Note. Two-way Student's t-tests of pre- minus post-test scores, thus positive values for the mean difference (and effect size d) indicate the presence of attrition, or a decrease in test scores; M = mean difference; SE = standard error of the mean difference; d = Cohen's d.

Figure 3. Distributions for the differences between pre- and post-test scores for both item formats.

Despite everyone being encouraged to do so, only some students studied with Memrise (n = 41) and Xreading (n = 60) over the summer break (see Table 3 and Figure
4 for descriptive statistics). Furthermore, only 11 students were classified as having travelled abroad to an English-speaking country. For each test format (recall and recognition), a multiple linear regression was calculated with JASP to predict the change in vocabulary scores based on minutes studied using Memrise, words read in Xreading, and whether the participant travelled abroad. The variables were added in order of predicted importance (Memrise first, followed by Xreading, and finally travelling abroad), but none were significant predictors of score changes, with only 2.3% of variance explained, $F(3, 98) = .78, p = .508$.

Table 3

<table>
<thead>
<tr>
<th>Treatment</th>
<th>n</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SEM</th>
<th>Median</th>
<th>SD</th>
<th>Skew</th>
<th>SES</th>
<th>Kurt</th>
<th>SEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memrise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>41</td>
<td>2</td>
<td>12,740</td>
<td>494.5</td>
<td>311.9</td>
<td>40</td>
<td>1,997</td>
<td>6.06</td>
<td>0.37</td>
<td>37.82</td>
<td>0.72</td>
</tr>
<tr>
<td>Xreading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>60</td>
<td>181</td>
<td>148,900</td>
<td>35,600</td>
<td>4,780</td>
<td>19,990</td>
<td>37,030</td>
<td>1.25</td>
<td>0.31</td>
<td>1.24</td>
<td>0.61</td>
</tr>
</tbody>
</table>

*Note. SES = standard error of skewness; SEK = standard error of kurtosis.*

Figure 4. Boxplots showing Memrise and Xreading use over the summer break (one extreme outlier who practiced for 12,738 minutes was removed from the Memrise graph and the multiple linear regression analysis).

Conclusions

Investigating RQ1, we were unable to find evidence of attrition in receptive vocabulary knowledge after two months of summer vacation. Rather, a significant, albeit small ($d = -.31$; Plonsky & Oswald, 2014), gain in meaning recognition scores was found. Furthermore, no meaningful relationships were found between individual differences in
pre- and post-test receptive vocabulary scores and the number of hours spent studying vocabulary through Memrise, the number of words read through Xreading, or experience travelling abroad over the summer holiday period (RQ2). Overall, these results support the arguments of some researchers that receptive vocabulary knowledge is not easily lost relative to productive knowledge (Schmitt, 2010; Weltens & Grendel, 1993).

**Future Directions**

Future research in this area should utilize productive measures of vocabulary knowledge in addition to other, more sensitive, measures of receptive vocabulary knowledge such as reaction time experiments or tests of collocational knowledge (as advocated by Weltens & Grendel, 1993). Furthermore, ensuring greater compliance with the recommended summer studying would perhaps aid statistical analyses by providing greater variance.

**References**


