Mobile learning of vocabulary from reading novels: a comparison of three modes

Tony Fisher, Richard Pemberton, Mike Sharples

Hiroaki Ogata, Noriko Uosaki

Philip Edmonds, Anthony Hull, Patrick Tschorn

ABSTRACT
A controlled study of vocabulary learning from reading compared three types of mobile technology: paper books, e-books with dictionaries, and e-books with adaptive software for vocabulary learning. The results showed small English vocabulary gains among Japanese students for all three conditions, with no significant difference between the technologies. The students reported that they found the paper book easy to use and annotate, but that the mobile devices enabled them to quickly look up the meanings of words. The additional functions of the adaptive device were only used by a few students. Recommendations are made for further research and design.

INTRODUCTION
Mobile technology offers a promising way to support language learning, through easy access to multimedia teaching materials from a portable device that can both provide teaching during the gaps in daily life and tools to support learning on demand outside the classroom (Ogata & Yano, 2003; Ogata, et al., 2006). Several recent studies of second language (L2) vocabulary learning through mobile devices have focused on instructed vocabulary learning, with SMS lessons ‘delivering’ vocabulary for learners to study at regular frequencies (e.g. Joseph, Binsted & Suthers, 2005; Song & Fox, 2005; Thornton & Houser, 2005; Kennedy & Levy, 2008; Lu, 2008). Studies of incidental mobile vocabulary learning, where the learner acquires vocabulary without consciously focusing on it, have so far been rare (one exception being Song & Fox, 2008). Similarly, although e-book reading on mobile devices is currently gaining popularity, there has been little examination of the potential of these technologies for supporting L2 vocabulary acquisition.

In this paper we report on a study that aimed to investigate incidental learning of L2 vocabulary from reading books. The study involved a carefully controlled comparison of learning from reading in three conditions: paper book, e-book with online dictionary, and e-book with adaptive vocabulary learning support – the English Language Mobile (ELMO) system.

The research was carried out for Sharp Laboratories of Europe. The Sharp team was responsible for the software design and log analysis; the University of Tokushima team was responsible for conducting the study; and the University of Nottingham team for designing and evaluating the study. Although the evaluation was funded by Sharp Laboratories of Europe and the research design was agreed in advance by all three parties, the conduct of the evaluation and reporting of results was not subject to any company influence.

MATERIALS
The paper books were graded readers from the Oxford Bookworms Library, published by Oxford University Press (OUP). The two e-book applications were custom-designed and developed by the company, and ran on Sharp

1 The Oxford Bookworms Library readers and dictionary were used by kind permission of Oxford University Press.
smartphones. As can be seen from the example screens in Figure 1, the applications displayed the text of the OUP books and provided an electronic English dictionary (the *Oxford Essential Dictionary*) for students to look up the meaning of a word by a click (some words had Japanese glosses). All interaction was through touch or stylus. The ELMO system (Figure 1b), also developed by the company, provided additional interactivity, including highlighting words in the text that matched the user’s reading level, short vocabulary learning activities accessible by a click (tab 2 in Figure 1b), and learning progress charts (tabs 3-4). The initial level of reading ability was assessed by the user taking a short vocabulary test when first using ELMO. Thereafter, the system employed an algorithm to adapt its response based on the user’s progress through the text, the words clicked, and activity success rate.

Two different smartphones were used in the trial, the Sharp WS004SH (3.7 inch colour TFT screen, 480x640 resolution) and the Sharp WS007SH (2.8 inch colour TFT screen, 480x640 resolution). Note that although both types of phone had identical resolution, the WS004SH had a larger -sized display. Although comparison of phones was not part of the original design, the two different phone types provoked some useful comparative responses.

METHODS

The study was conducted at a Japanese high school over a six-week period and involved 39 Year 10 and 11 students, aged 15 to 17. A ‘crossover’ research design was employed, involving three comparison groups of 13 students each, counterbalanced for order effects of technology. The groups were matched for score on a pre-test of English vocabulary (Paul Nation’s Vocabulary Levels Test). After discussion with the research team, the students’ form teacher chose three readers from the Oxford Bookworms Library for use in the study: *The Hound of the Baskervilles*, *The Thirty-Nine Steps*, and *Little Women*. All books were Stage 4 Bookworms and had a restricted vocabulary of 1,500 headwords2 or word families. Each novel was made available in each of the three conditions: paper book, a basic e-book with online dictionary, and the adaptive ELMO system. Each group of students engaged with all three conditions and the order of the technology conditions was rotated, so that each group had a different start condition (see Table 1).

| Cycle A: *The Hound of the Baskervilles* | Cycle B: *The Thirty-Nine Steps* | Cycle C: *Little Women* |
| Wk1 | Wk2 | Post-test one | Wk3 | Wk4 | Post-test two | Wk5 | Wk6 | Post-test three | Survey and focus groups |
| GpA | Baseline test | ELMO | e-book | Wk3 | Wk4 | Post-test one | Wk5 | Wk6 | Post-test three | Survey and focus groups |
| GpB | Pre-test | e-book | book | | | | | | |
| GpC | | book | ELMO | | | | | | |

Table 1. Summary of the research design.

**Initial test**

The students underwent an initial test in two parts. The first was a 120-item standardised baseline test of English vocabulary (the Vocabulary Levels Test), testing knowledge at the 2,000, 3,000, 5,000 and 10,000 word frequency levels.

---

2 A headword is the word at the front of a dictionary entry. For example, *go, goes, going, gone* and *went* all belong to the headword ‘go’.
Its purpose was to enable allocation of the students to three sub-groups, matched by vocabulary level, and with a similar gender distribution.

The second part was a 90-item pre-test of words selected from the three OUP books (30 words/phrases per book, made up of 20 less common words/phrases from outside the 3,000 frequency range and 10 more common words/phrases from the 2,000 frequency range from each book). None of the words/phrases selected from each book occurred in the other two books. The purpose of the pre-test was to enable comparison with the post-test scores for each book and each condition.

**Intervention**

In each cycle, the novel was first issued to students in the appropriate format, according to Table 1 above. Students were asked to read the novel over the following two weeks, allocating as much time as they felt appropriate, and using whatever dictionary support they felt appropriate. It was stressed that whilst they were free to discuss the novel, they were not to swap devices or books with one another; and they were given a date two weeks later when they should return their book/device, and when they would sit a vocabulary test on the book they had read.

The post-test for each book, unbeknown to the students, was composed of the same 30 words which had been used in the 90-word pre-test, enabling a direct comparison to be made between the pre- and post-test scores for each book. The pre- and post- tests were marked by one of the Japanese researchers, each item being scored as either correct or incorrect. Students carried out the reading either at home or on the train journey to and from school.

The third cycle was followed by a two-part questionnaire survey and focus groups. Questionnaires were designed in English and were translated into Japanese by one of the Japanese researchers. Three balanced focus groups of 13 students each were formed, each containing students from each of Groups A, B and C. The focus groups were conducted in Japanese by the Japanese researchers using a schedule of probes and prompts.

**RESULTS**

The pre- to post-test differences between the mean test scores for novel and for technology are shown in Table 2, along with the mean standard deviations.

<table>
<thead>
<tr>
<th>Novel A</th>
<th>Mean difference pre-post</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hound)</td>
<td>1.92</td>
<td>3.10</td>
</tr>
<tr>
<td></td>
<td>1.39</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td>1.54</td>
<td>1.94</td>
</tr>
<tr>
<td>Novel B</td>
<td>Mean difference pre-post</td>
<td>1.85</td>
</tr>
<tr>
<td>(39 Steps)</td>
<td></td>
<td>5.27</td>
</tr>
<tr>
<td></td>
<td>1.10</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>1.10</td>
<td>4.65</td>
</tr>
<tr>
<td>Novel C</td>
<td>Mean difference pre-post</td>
<td>0.38</td>
</tr>
<tr>
<td>(Little Women)</td>
<td></td>
<td>2.47</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td>-1.15</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Table 2. Pre-post differences between means and standard deviations for the test scores.

A within-subjects ANOVA analysis of the test scores by technology (ELMO, e-book, book) and time (pre, post) showed a main effect of time \([F(1,38)=12.62, \ p<0.001, \ \eta^2 = .25]\) but no effect of technology and no interaction between technology and time \([F(1,39)=.72],\) as can be seen from the overlapping ranges of the standard error bars in Figure 2.
So students learned from the experience (overall mean pre-test score = 6.55 out of 30, overall mean post-test score = 7.5 out of 30) but the increase in scores was not associated with any particular technology. Moreover, this improvement only has a small effect size, so whilst the result is statistically significant, it may be considered as not particularly ‘educationally significant’.

**DISCUSSION**

The differences in the means shown in Table 2 indicate that students learned, on average, only one new word over each two-week period, regardless of the technology. One explanation for this poor performance is that they did not engage with any of the methods of learning. A study of the log files shows that few students read an entire book, with most reading three pages or fewer out of some 100 pages in each book. It would appear that the students found the books difficult and the teacher overestimated their English capability. They were also very busy with official homework and this additional reading was not a priority for them as it did not affect their all-important school reports at the end of the term.

Although differences in preferences for the three modes are also not significant, the quantitative and qualitative data do suggest certain tendencies. For example, the focus group interviews revealed that several students preferred the paper books for their portability, ease of use, lack of strain on the eyes, and the ability to annotate the text. On the other hand, looking up in a separate dictionary was felt to be “troublesome” (Focus group 3) and the devices were generally seen to be more effective for vocabulary learning:

“When you want to enjoy reading, ‘paper book’ is better and when you want to learn words, devices are better.” (Focus group 1)

“I just kept reading without knowing the meanings of some words thus hardly followed the story well [with the paper book]. On the other hand, the devices were more useful because I checked new words, thus I could follow the story more.” (Focus group 1)

Of the three modes, ELMO was rated as both the most helpful and the most enjoyable, according to the questionnaire data: “It was a little fun to do, so it motivated me to learn more” (Questionnaire B, ELMO). Given that several respondents commented that they did not know of the additional activities provided in ELMO, and few found or used the progress chart, it seems probable that reactions to ELMO would have been more positive if the features available in ELMO had been widely known and used.

However, although some students wanted ELMO functionality in their existing phones, they were not willing to purchase an extra device:

“I do not want to bring another device with me other than my mobile phone to read novels” (Focus group 1)

“Many say that if it is possible to use ELMO by inserting SD card to their own mobile phones, they would recommend it to their friends.” (Japanese Evaluator, Focus group 1)

**CONCLUSIONS**

Adaptive software linked to e-books running on mobile devices could offer an engaging means for students to improve their knowledge of a foreign language through incidental learning. But a number of factors need to be considered when
designing such systems. Ideally, the software should run on devices that the students already own. It should complement the reading of the book, rather than detract from it, and it should offer facilities for annotating the text. One possibility would be for students to first read for pleasure, marking troublesome words, and then later return to this list for self-study. Another would be to consider game-like opportunities where some demonstration of learning or problem-solving might be required for the user to move onto the next book/level. In terms of research design, care also needs to be taken to ensure that participants can spend sufficient time on task for incidental learning gains to occur. We are conducting further research that takes these factors into account.

ACKNOWLEDGMENTS
The authors wish to thank the students and their teachers who participated in the study for their contributions and constructive comments.

REFERENCES
Song, Y. and Fox, R. Using PDA for undergraduate student incidental vocabulary testing. ReCALL 20 (2008), 290-314.