Incidental second language vocabulary learning from reading novels: a comparison of three mobile modes

Tony Fisher, University of Nottingham
Mike Sharples, The Open University UK
Richard Pemberton, University of Nottingham
Hiroaki Ogata, University of Tokushima
Noriko Uoasaki, University of Tokushima
Phil Edmonds, Anthony Hull, Patrick Tschorn, Sharp Labs Europe

Abstract
This paper reports on a study in which incidental English vocabulary learning from three mobile modes (book, e-book and e-book with user modelling and adaptive vocabulary learning support) was investigated. The study employed a crossover design to test for vocabulary gain from reading three simplified English novels among a group of Japanese high school students, learning English as a second language. Small vocabulary gains were noted; however there was no significant difference between the modes in this respect. Participants also gave their reactions to using the three modes. We reflect on some possible reasons for the results, and identify some methodological considerations.

Keywords: mobile assisted language learning (MALL), vocabulary learning, incidental learning, extensive reading, adaptive systems, electronic dictionaries, second language, L2

Dedication
The authors of this paper wish to dedicate it to the memory of our friend and colleague Richard Pemberton, who died during the later stages of its preparation. Richard was a fine teacher, scholar and researcher and is missed by those who knew and worked with him.

INTRODUCTION
The acquisition of vocabulary is an essential aspect of learning a language. Research conducted in America (Anderson et al., 1988) found that the median 5\textsuperscript{th}-grade primary school student read approximately 600,000 words of running text in out-of-class reading per year (with exposure ranging from 0 to four million words); based on this finding it is estimated that including in-class reading, on average school students in English-speaking countries read approximately one million words of text per year (Nagy & Anderson, 1984; Nagy et al., 1985; Nagy, 1997). However, learners of English in non-English-speaking countries generally encounter English only as a subject of study and are therefore not exposed to anything like the amount of English text as their counterparts in English-medium education. Given the importance of vocabulary knowledge for language learning, a major issue for learners of English as a foreign language (EFL) is how to increase vocabulary size to a point at which they can read English texts fluently and extensively.

There are two approaches to vocabulary acquisition in EFL learning: direct instruction with memorisation, and incidental learning. Direct instruction is generally recommended in the second language vocabulary acquisition literature for learning the most common words of
English that occur frequently across a wide range of texts, whereas incidental learning is recommended to build up a vocabulary that goes beyond the 2,000 or 3,000 most common words of English (e.g. Nation, 2001). This however brings us to the problem of time. If English is taught for only a few hours per week, and if much of this time is spent on listening to the teacher, doing written exercises, or engaging in group work, then how can students progress beyond a basic level of reading vocabulary? Teachers typically encourage their students to access English materials out of class in web or paper form, but the vast majority of EFL school students find authentic written text too difficult to handle. Simplified and graded English novels in book form offer learners a chance to read widely at a level that suits them but may not be convenient for the mobile generation, and typically do not provide glosses to explain unknown vocabulary.

Here we report an investigation into the effectiveness of incidental vocabulary learning among Japanese high school students. The study employed simplified novels in three different modes:

- A paper book
- A mobile phone e-book incorporating an online dictionary
- A mobile phone e-book with enhanced software providing adaptive vocabulary learning support: the English Language Mobile (ELMO) system.

The central research question was, ‘Is the mobile mode in which a novel is read, associated with differential ‘incidental’ English vocabulary learning among a group of Japanese school students?’ This paper focuses particularly on some of the methodological issues encountered in setting up and running such a study.

**RESEARCHING LANGUAGE LEARNING AND MOBILE DEVICES**

This study draws on two distinct strands of research, one being incidental second language vocabulary acquisition, and the other being language learning with mobile technologies.

In the learning of any language, the progressive learning of vocabulary plays an essential part. In the past, learning second/foreign language vocabulary lists by rote was often a favoured approach (intentional learning). This has, to varying extents, been supplanted by more communicative and meaning-focused approaches, according to which vocabulary is acquired while the learner is focused on communicative tasks (incidental learning). Two key proponents of incidental learning have been, in the field of second language acquisition, Krashen, who argued that vocabulary is acquired through extensive reading (e.g. Krashen, 1989); and in the field of first language acquisition, Nagy and associates, who argued that students learn far more words incidentally through reading than is possible through direct instruction (e.g. Nagy et al., 1987).

Based on the figure that we mentioned earlier of exposure to approximately one million words of written text each year, Nagy and associates (Nagy et al., 1985; Nagy et al., 1987; Nagy, 1997) have estimated that school students in the USA are likely not to know the meaning of 2% of the words they read (20,000 words), are likely to learn 5% of these unknown words incidentally, and are therefore likely to learn some 1,000 words incidentally from their reading each year. However this figure, based on the research of Nagy et al. (1987), involving a delayed post-test, may be conservative. In a meta-analysis of incidental ‘mother tongue’ (first language: L1) vocabulary learning through reading, Swanborn and de Glopper (1999) investigated twenty experiments and concluded that “under natural reading
circumstances students will spontaneously derive and learn the meaning of about 15 words on every 100 unknown words they encounter” (p. 279). This would mean that if they encounter 20,000 unknown words a year, they would acquire 3,000 of these through incidental learning, or an average of eight words per day.

Swanborn and de Glopper, however, largely focused on immediate post-test studies, and specifically excluded studies of second language (L2) learning from their analysis. Huckin and Coady (1999) on the other hand, focused specifically on studies of incidental L2 vocabulary acquisition, concluding that such acquisition is important as it is ‘the primary means by which second language learners develop their vocabulary beyond the first few thousand most-common words’ (Huckin & Coady, 1999, p. 190) and that a basic sight-recognition vocabulary of at least 3,000 word families is a pre-requisite for such incidental learning. In both cases (L1 and L2), the incidental learning of additional vocabulary occurs as a result of the readers’ attempts to get meaning from what they are reading - normally a fictional narrative.

With wide ownership of mobile devices, the potential of using such devices to support language learning has become evident. Such learning is facilitated by easy access to multimedia materials from a portable device that can provide a learning opportunity during leisure time (Kadyte, 2004) and in the gaps between other activities in daily life, through an increasingly wide range of tools to support learning on demand outside the classroom (Ogata & Yano, 2003; Ogata et al., 2006). As Kukulska-Hulme and Shield (2008) noted in their review paper of developments in ‘mobile assisted language learning’ (MALL), the range of approaches and learning activities using MALL is developing very quickly, expanding in the space of two or three years from a purely teacher-learner, text-based model to one that is beginning to support multimedia, collaborative listening and speaking activities and to allow learners to co-construct knowledge to solve problems and fill information gaps. (p. 283)

Several recent studies of MALL have focused on vocabulary learning. The majority of these have aimed to aid intentional vocabulary learning: for example, by pushing vocabulary items for learners to study at regular frequencies through email or SMS (e.g. Song & Fox, 2005; Thornton & Houser, 2005; Kennedy & Levy, 2008; Lu, 2008); by recommending vocabulary for learning according to the learner’s previous memory cycles (e.g. Chen & Chung 2008); by providing contextually-relevant vocabulary for learning or use (e.g. Ogata & Yano, 2005; Beaudin et al., 2007; Edge et al., 2011) or by testing knowledge of objects within a particular environment (Ogata et al., 2010); and by enabling users to develop collaborative knowledge of vocabulary items by sharing photos, videos, definitions, example sentences, tags etc. (e.g. Joseph et al., 2005; Pemberton et al., 2010; Wong & Looi, 2010).

However, studies of incidental mobile vocabulary learning have been less common. Song and Fox (2008) investigated three Hong Kong students’ use of e-dictionaries on a phone/PDA to support their academic study in an English-medium university when, for example, encountering unknown words in lectures or whilst reading. Thus the vocabulary learning was incidental, but the texts were not presented on mobile devices. The closest study to our own is that of Watson, Todd and Tepsuriwong (2008), who developed English reading mazes (branching stories) for mobile phones from which Thai readers reported vocabulary gains. However, their application was designed as a ‘game’ to enhance motivation and featured branching stories with one or two sentences per screen rather than a typical e-book content or format.
Although the reading of e-books on mobile devices is currently gaining in popularity, we are not aware of any prior research into the potential of these technologies for supporting incidental L2 vocabulary acquisition.

**METHOD**

Here we report the results of a recent study (conducted in 2009) that aimed to investigate incidental learning of L2 vocabulary by Japanese high-school students engaged in reading English novels in informal settings. The study involved a 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.

Three distinct teams of researchers were involved in the research, which was commissioned by Sharp Laboratories of Europe: a team at Sharp was responsible for the software design, the provision of mobile devices, and the processing of data from device logs; a team from the Learning Sciences Research Institute at the University of Nottingham undertook the design, evaluation and reporting of the study; a team from the University of Tokushima was responsible for conducting the study in the field, including briefing the students, distribution and collection of devices, and data collection from the participants. Although the evaluation was funded by Sharp Laboratories of Europe and the research design was agreed in advance by all three parties, neither the conduct of the evaluation nor the reporting of results was subject to any company influence.

**Participants**

The University of Tokushima team located a high school in which a class teacher was willing for one of her classes to participate in the study. Finding a willing teacher was not easy, as the Japanese high school system is subject to many pressures; further, although a ‘pen drive’ was given to each student participant as a ‘thank-you’ for participation, nothing could be accepted by teachers or the participating school. Students’ informed consent was gained for their participation in the study. Students were also given a letter of information for their parents.

The study group selected by the teacher comprised 39 students from Years 10 and 11 (24 female, 15 male), aged 15 to 17. The students were judged by their teacher to be working at Level 4 of the Oxford Bookworms Library of English readers, published by Oxford University Press.

Each student had a unique student identity number, the key to which was known only to the class teacher who was not otherwise involved in the research. This number was used throughout the study on all tests and other instruments, in order to protect student anonymity.

**Research design**

For this exploratory study, a ‘crossover’ research design was developed, involving three comparison groups of 13 students each, counterbalanced for order effects of technology. A crossover design is a longitudinal study in which participants engage in a sequence of interventions, such that each person participates in all conditions.

The evaluation was carried out over a period of six weeks. Each group of students engaged for two weeks with each of the three learning modes, as indicated in Table 1. The order of the
technology conditions was rotated, so that each group had a different start condition. The sequence of novels and technologies was the same for all three groups, so the study tested three out of six possible orderings.

The groups were matched for score on a baseline pre-test of English vocabulary. All students sat the baseline test, involving words from the three readers. These words would, unbeknown to the students, also appear as the respective post-test for each novel, at the end of each two-week cycle. Testing the same vocabulary for each condition has the advantage of comparing identical test items, but the possible confounding factor of learning through repeated testing. The results suggest that this did not occur. At the change-over between cycles, the phone SD cards would be retrieved from the devices to enable safe-keeping of the log data contained on them.

At the conclusion of all three cycles, attitudinal data were collected from the participants by means of questionnaires and focus group discussions.

<table>
<thead>
<tr>
<th>GpA</th>
<th>Base-line test</th>
<th>Cycle A - The Hound of the Baskervilles</th>
<th>Cycle B - The Thirty-nine Steps</th>
<th>Cycle C - Little Women</th>
<th>Survey and focus groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wk1</td>
<td>Wk2</td>
<td>Wk3</td>
<td>Wk4</td>
</tr>
</tbody>
</table>

Table 1: Summary of the evaluation design

This design was adopted to enable assessment of the relative impact of each of the three modes. The advantages of a crossover design for this study are that since each group has a different start condition, the order effects of technology type are removed (each technology is tested at first, second and third position in the sequence). Also, each participant experiences all three types of technology, so can compare subjective experiences.

With a relatively small sample, such as we had for this study, a between-groups design would have had much less power to find any effects. A crossover design provides a greater possibility of revealing a small to medium effect size, as although each of the three conditions has 13 participants at any one time, all 39 participants experience each condition during the study.

The main limitation of a partial crossover design is that the sequence of treatments may affect the outcomes. Although each group started with a different technology, they all progressed in the same order (e.g. e-book followed ELMO). However, there was no reason to expect that using one particular technology would influence the learning activity of the next and the results suggest that this was not the case.
Materials
After discussion with the research team regarding suitable texts and levels, the teacher chose the three novels used in the study (The Hound of the Baskervilles; The Thirty-nine Steps; Little Women) from the Bookworms Level 4 range of available titles. Each novel has a restricted vocabulary of 1,400 headwords or word families. Whilst this is lower than the threshold working knowledge of 3000 proposed by Huckin and Coady (1999) as necessary for incidental learning to occur, the higher threshold relates to unsimplified text; in our study simplified texts were used, so incidental learning remains feasible in this context. 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.

The paper books are graded readers from the current Oxford Bookworms Library. The two e-book applications are custom-designed and developed by Sharp, and run on Sharp smartphones. As can be seen from the example screens in Figure 1, the applications display the text of the OUP books and provide 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 with the device is through touch or stylus.

The ELMO system (Figure 1b), also developed by the company, provides additional interactivity, including the highlighting of words in the text that matches 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). In the case of the devices running the ELMO system, a user’s initial level of reading ability is assessed on first start-up by the user taking a short vocabulary test to provide the basis of a user model. Thereafter, the system employs 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, being the Sharp WS004SH and the Sharp WS007SH. Although both devices had identical screen resolution (480x640), the WS004SH had a larger-sized display (3.7 inch colour TFT screen) than the WS007SH (2.8 inch colour TFT screen). Although the comparison of device screen size was not part of the original research design, the two different device types provoked some useful comparative responses from users during the final survey and focus groups.
**Procedure**

The students underwent an initial test in two parts. The first part was a 120-item standardised baseline test of English vocabulary (Paul Nation’s Vocabulary Levels Test\(^3\)), testing knowledge of vocabulary at the 2,000, 3,000, 5,000 and 10,000 word frequency levels. This test was conducted in order to enable a balanced allocation of the students to three subgroups, matched by vocabulary level of group members, and with a similar gender distribution. For this reason it was necessary for this test to be conducted several days before the first cycle of reading, in order to allow time for the test to be marked and for the groups to be constructed.

The second part of the initial test was a 90-item pre-test of words selected from the three OUP books. For each book a set of 30 words/phrases was selected, 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. None of the words/phrases selected from each book occurred in the other two books. The purpose of this pre-test was to enable comparison with the post-test scores for each book and each condition. Students were asked to respond to the given English words/phrases by giving either a Japanese word with the same meaning, a definition in Japanese, or a definition in English. If they did not know a particular word/phrase, they were asked to confirm this by ticking a box in order to confirm that no questions were simply skipped, either deliberately or inadvertently. The pre- and post- tests were marked by one of the Japanese researchers (who is also an experienced EFL teacher), each item being scored as either correct or incorrect.

Students then engaged with the novels. In each cycle, the novel was first issued to students in the appropriate format, according to Table 1 (above). Those using a device were also given a help sheet in case of need. Students were asked to read the novel in their own time over the following two weeks, allocating as much time as they felt appropriate, and using whatever dictionary support they needed in order to assist their understanding. It was stressed that, whilst they were free to discuss the novel, they were not to swap devices or books with one another. They were given a date two weeks later when they should return their book/device, and when they would sit a vocabulary test based on the book they had read. Whilst this last point is not entirely consistent with other studies in incidental learning, it was necessary from the point of view of students’ informed consent. Students were given no guidance as to the content of the test and were not asked to prepare for it in any way. It was also emphasised that their performance in this test would not affect their class position.

Students carried out the reading wherever they wished, for instance at home or on the train journey to and from school. 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. In each case the vocabulary test was carried out after school, under test conditions. Whilst students were completing the test, the local researchers removed and replaced the SD cards.

The third cycle of reading was followed by a two-part paper-based questionnaire survey and focus groups. Questionnaires, comprising a mixture of statements with Likert scale responses and open questions with free-text responses, were designed in English and were translated into Japanese locally by one of the Japanese researchers. Three balanced focus groups of 13 students were formed, each containing students from study Groups A, B and C. The focus groups were recorded on video and were conducted in Japanese by the Japanese researchers using an agreed schedule of probes and prompts.
RESULTS
Analysis of the test results was undertaken using SPSS. The pre- to post-test differences between the mean test scores for novel and for technology are shown in Table 2, along with the standard deviations for each set of results. There is no statistically significant difference between the three conditions of ELMO, e-book or book, nor interactions between technology and novel.

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

Table 2. Descriptive statistics for pre-post test differences by novel and by technology

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, η² =.25] but no effect of technology and no interaction between technology and time [F(1,39)=.72]. So whilst 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) 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’.

Hence, in this study, apparent differences between technologies are not meaningful, as can be seen from the overlapping ranges of the standard error bars in Figure 2.

Figure 2. Pre-post mean vocabulary test scores by technology, standard error bars added
The differences in the mean scores shown in Table 2 indicate that students learned, on average, one new word over each two-week period, irrespective of which technology was used. One explanation for this generally poor performance is that they did not engage greatly with any of the modes of reading, nor with ‘the story’ of any of the novels. 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. Similarly, from the questionnaires, few students (5) claimed to have finished one or more of the novels.

It would appear that the students found the books difficult to read and that their teacher may have overestimated their English capability. This view is supported by the low mean pre-test scores for each novel, as can be seen in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test mean</th>
<th>Pre-test SD</th>
<th>Post-test mean</th>
<th>Post-test SD</th>
<th>Pre-post mean difference</th>
<th>Pre-post SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hound of the Baskervilles</td>
<td>4.10</td>
<td>2.12</td>
<td>5.72</td>
<td>3.40</td>
<td>1.62</td>
<td>2.42</td>
</tr>
<tr>
<td>Thirty-nine Steps</td>
<td>7.10</td>
<td>4.33</td>
<td>8.44</td>
<td>4.49</td>
<td>1.33</td>
<td>4.23</td>
</tr>
<tr>
<td>Little Women</td>
<td>8.50</td>
<td>3.13</td>
<td>8.36</td>
<td>3.34</td>
<td>-0.10</td>
<td>2.28</td>
</tr>
</tbody>
</table>

Table 3: Descriptive statistics for pre and post-test results by novel

Though the novels were all from the same level (‘level 4’), in practice the vocabulary posed differing levels of demand, as may be seen from the pre-test results in Table 2. Also of note in Table 2 are the generally large standard deviations (SD), being a consequence of the many ‘outliers’ among the scores.

Some important findings from the questionnaires and focus groups were as follows:
- There was no significant difference in expressed preference between the three modes
- The focus groups revealed that some students preferred paper books
- Students liked not having to carry additional dictionaries
- Students were attracted by the technology
- ELMO was rated as the most helpful and enjoyable of the three modes
- Students expressed an unwillingness to carry multiple devices

Although differences in preferences for the three modes are not statistically significant, the quantitative and qualitative data do suggest certain tendencies. Despite the assumptions which are often made about young people and digital screen-based technologies, the focus group interviews revealed that several students preferred the paper books for their portability, ease of use, lack of strain on the eyes, the ability to annotate the text, and ‘paper touch [which] makes me feel I’m reading’. On the other hand, looking up words in a separate dictionary (which also had to be carried around if the paper books were to be used on the move) 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 could not look up, since I did not have dictionary with me while reading in the train. (Questionnaire B)

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)

From the questionnaire responses it was evident that some students were attracted by the technology of the devices: ‘looks cool, stylish’, and by the informality of use: ‘able to read casually, and I liked touching by pen (stylus)’; ‘able to do as if I'm playing a game’; ‘able to read where it's dark’ (Questionnaire B).

<table>
<thead>
<tr>
<th></th>
<th>Paper book</th>
<th>e-book</th>
<th>ELMO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enjoyability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean score /5</td>
<td>3.10</td>
<td>3.23</td>
<td>3.45</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.82</td>
<td>0.87</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Helpfulness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean score /5</td>
<td>2.82</td>
<td>3.10</td>
<td>3.26</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.97</td>
<td>0.94</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics for enjoyability and perceived helpfulness in ‘growing your English vocabulary’ by mode

Though the differences shown in Table 4 are not large, 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)

Although some students wanted ELMO functionality in their existing phones, they appeared unwilling to purchase or carry 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)

**REFLECTIONS ON THE METHODOLOGY AND LESSONS LEARNED**

This study was carefully planned, and was interesting and challenging to conduct. However, the incidental gain in vocabulary was not as great as we might have expected to see, irrespective of mode. This raises the obvious question: why have some previous studies of language learning with mobile devices reported clear benefits, whereas this one has shown minimal gains? We suggest that this could be for a combination of four reasons.

First, other studies may be looking at different aspects of language learning, such as reading comprehension (e.g. Chen & Hsu, 2008) or communicative use of language (e.g. Watson Todd & Tepsuriwong, 2009). Second, they may use different measures of success such as variety of usage (Song & Fox, 2008) or motivation to learn (e.g. Wong & Looi, 2010). While
motivation can be a condition for successful learning, it is not a measure of the learning. The participants may be motivated by engaging with other marginal or distracting activities, such as exploring the technology or playing games with the system (Baker et al., 2004). Third, other studies may have a similar design to the one reported here, but with different participants (e.g. Edge et al., 2011; Sandberg, Maris & de Geus, 2011), setting, teaching materials, or technology, such as sending vocabulary items by SMS to students’ phones (e.g. Thornton & Houser, 2005). Fourth, the studies may focus on intentional, rather than incidental, language learning (e.g. Kennedy & Levy, 2008; Lu, 2008; Cavus & Ibrahim, 2009).

We have not found any other studies that have used the combination of participants and design reported here. From our results, it is clear that the particular combination of participants (Japanese school students aged 15-17), setting (incidental learning out of school), teaching materials (classic novels) and technology (vocabulary learning from books or smartphones) was not conducive to incidental vocabulary learning. Some possible reasons for this are given below, as a guide to future development of incidental vocabulary learning on smartphones.

The Japanese school students participating in the study do not appear to have had sufficient ‘spare time’ to engage fully with the reading. We can conjecture that if they had been in a position to spend more time reading and engaging with associated activity, then more incidental learning might have occurred, bringing with it a greater possibility of differences emerging between the modes.

The participants, who were seen as good, conscientious students who take their schoolwork very seriously, were very busy with official school homework; hence this additional reading was not a priority for them, particularly since it would not affect their all-important school reports at the end of the term. Any presumed motivational effect of reading a novel in a language which they were studying as a school subject, or of using a new device, was insufficient to overcome the twin constraints of language difficulty and pressure of time. Hence, it would be important in any similar studies in the future to ensure that participants can spend sufficient time on task for incidental learning gains to occur.

One particular time-related difficulty encountered with the implementation of the research design was that of finding a run of six weeks during which school commitments for the students would be broadly consistent, and during which students would be available to exchange devices and undertake the post-tests on the appropriate dates, whilst also avoiding holidays, school trips and other such scheduled events.

The books were chosen by the class teacher from a constrained list of available titles (for commercial reasons, only classic books out of copyright were available for the study). It is possible that some other titles, perhaps nearer to the experience and interests of Japanese teenagers, might have proved to be more motivating. The teacher initially involved was herself extremely busy, and left the school to take up another post part-way through the project.

Other studies of incidental vocabulary gain from EFL reading among Japanese high school students are rare, although studies at university level have tended to find similarly low rates of long-term incidental vocabulary learning from reading (Waring & Takaki, 2003; Brown et al., 2008). The gain of around one word per two-week cycle in our study was similar to that
among Japanese high school students in an earlier study conducted by Day et al. (1991). However, in contrast to our study, which involved a pre/post methodology, theirs involved a ‘treatment/control’ approach and was conducted over a much shorter time-scale. The treatment group read an adapted short story in the classroom during about 30 minutes of ‘sustained silent reading’. This was followed immediately by a 17-item multiple choice test in which the treatment group outperformed the control group by an average of one word. By definition a multiple choice test provides the correct answer among the choice offered, whereas, by only providing the English term to be defined, the test we used was more challenging and included a greater number of items (30).

In relation to time-scale, Day et al. note that many of the words would not be retained in the longer term owing to the limited context in which they had been encountered. In our study we assumed that participants would need two weeks to read each of the novels involved in their spare time, so even if they had been able to spend sufficient time on the reading involved, any incidental learning of specific vocabulary would have had to have been retained over a longer period in order to be available to be drawn on during the tests.

A further limitation was that there was not enough ‘repetition’ of vocabulary in the three texts used in our study to make incidental learning likely and this is something that would need to be addressed in future. Hence our study confirms the findings of previous studies which have found low gain rates where words have not been repeated. For example, chances of learning an L2 word incidentally have been found to be around 33% with eight encounters with the word (Horst et al., 1998; Waring & Takaki, 2003), but as high as 60% with 20 encounters (Pigada & Schmitt, 2006). In other words, research suggests that at least eight encounters are needed to provide a reasonable chance of incidental learning, but 20 encounters would be ideal. If words are to be encountered infrequently, perhaps just once, then a great deal of reading must be undertaken for any incidental learning to lead to significant gains.

In terms of their use of the technology, it became clear during analysis that there were problems with interface and interactivity. Some participants in our study did not log out reliably. This led to long, and sometimes apparently unfinished, sessions in the log files. An arbitrary period of ten minutes of inactivity was applied in order more reliably to demarcate sessions of activity. Participants were left to discover any additional functionality on the devices for themselves. 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. Hence some sort of briefing or familiarisation may have been appropriate, despite our intention not to appear to favour or emphasise any mode over the others. In summary, the activity of incidental vocabulary learning from classic English texts was not sufficiently appealing, in any of the conditions, for hard-pressed Japanese students to engage in study that was not assessed for their school exams.

CONCLUSION

The qualitative results gained from the questionnaires and focus groups have been the most helpful outcome in directing future development of the ELMO application. Sharp Laboratories of Europe is now emphasising features that will engage and motivate learners to continue using the application over longer periods of time. Different types of learner will be
motivated in different ways, but it is likely that development will address at least two core demographics: the formal student population (e.g., high school students as in the study) and the informal post-school adult learner. In the former case, motivation would be derived from a close integration with curriculum content, classroom methods, and assessment. In the latter case, though not specifically raised by the study, a wide range of content is required to satisfy different learner interests in order that reading English as a foreign language can become a pleasure or even a leisure activity.

Adaptive software linked to e-books running on mobile devices could offer an engaging and effective means for students to improve their knowledge of a foreign language through incidental learning. However, our research question here was, ‘Is the mobile mode in which a novel is read, associated with differential ‘incidental’ English vocabulary learning among a group of Japanese school students?’ The answer to that question would be ‘no’ on the basis of this study; nonetheless, this does not mean that the potential is not there. We therefore suggest that further investigation, employing the research design elaborated here and bearing in mind some of the ‘lessons learned’, is merited.

Acknowledgements
The authors wish to thank the students and their teachers who participated in the study for their contributions and constructive comments, and Oxford University Press for their assistance. Dr. Shaaron Ainsworth (University of Nottingham) provided important advice during the course of this study. The study was made possible by funding from Sharp Laboratories Europe. A shorter version of this paper appeared previously in the Conference Proceedings of mLearn 2009.

References


Kukulska-Hulme, A. & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL* 20(3), 271-289.


1 The Oxford Bookworms Library readers and dictionary were used by kind permission of Oxford University Press.

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’.