

Enhancement of automatization through vocabulary learning using CALL: Can prompt language processing lead to better comprehension in L2 reading?

TAKESHI SATO

*Tokyo University of Agriculture and Technology, Japan
(email: tsato@cc.tuat.ac.jp)*

MITSUYASU MATSUNUMA

*Shizuoka University, Japan
(email: ommatun@ipc.shizuoka.ac.jp)*

AKIO SUZUKI

*Toyo University, Japan
(email: suzukiakio@toyo.jp)*

Abstract

Our study aims to optimize a multimedia application for vocabulary learning for English as a Foreign Language (EFL). Our study is based on the concept that difficulty in reading a text in a second language is due to the need for more working memory for word decoding skills, although the working memory must also be used for text comprehension skills. This implies that the automatization of word decoding to free the working memory is indispensable. Therefore, developing a multimedia application for vocabulary learning with a time-control function, our study hypothesizes that vocabulary learning with this function brings about the automatization of word decoding or faster recall of the meanings, and, as a result of freeing more working memory, better comprehension of text. According to our research, multimedia application users not only retained more words and recalled their meanings more quickly but also obtained higher scores on reading comprehension tests that contained the words learned through the application. From this result, we conclude that multimedia vocabulary learning with a time-control function can automatize word decoding skills and therefore free more working memory, thereby leading to better comprehension of the text.

Keywords: L2 reading, working memory, automatization, vocabulary learning, time-control function, text comprehension.

1. Difficulty in comprehending texts

Reading activity, whether performed in our first language (L1) or second language (L2), is a higher cognitive process by its nature (Nation & Angell, 2006; Stuart, 2006).

This is due to the reading activity comprising multiple processes. In Harrington and Sawyer (1992), the reading process includes both a bottom-up process and top-down process, referring to decoding skills such as character and word recognition and comprehension skills such as an inference, respectively. In a reading activity, these processes arise spontaneously and interact with each other (Adams & Collins, 1979), and effective reading can be accomplished with both decoding and comprehension skills (Gough & Tumber, 1986).

In order to conduct such multicomponent cognitive processes, working memory (WM) plays a pivotal role. WM can be defined as mental processes that include the temporary storage and manipulation of information during on-going processing (Juffs & Harrington, 2011). This means that WM is not just a part of memory, but rather a part of cognition (*ibid.*). As WM involves the storage and processing information, its functions are crucial, particularly in performing activities such as comprehension, learning and reasoning (Baddeley, 1986), all of which are related to the reading process. Most skills leading to better reading comprehension such as inference, comprehension monitoring, and context structure comprehension, need WM for storage, retrieval, and integration of information (Cain & Oakhill, 2007). In order to examine the relationship between WM and reading comprehension, many studies focused on the difference between skilled readers and less skilled readers (for example, Oakhill, Cain, & Yuill, 1998; Perfetti & Lesgold, 1977).

However, information storage capacity in WM is limited. The limited information in WM is available for only a few seconds and then is lost. There is a trade-off between the processing and retaining functions in order to decide the amount of cognitive resources distributed to each function (Carpenter & Just, 1989). Therefore effective reading cannot be accomplished if too much WM is used for decoding skills such as word recognition, since proportionately less WM is available for information processing, which is the primary use of WM among skilled readers.

Furthermore, L2 reading tends to present more difficulties than L1 reading does. Although WM capacity for processing L2 is smaller than that for L1 (Berquist 1997), L2 readers need more WM, especially for decoding. Balass, Nelson and Perfetti (2010) argue that difficulties in text comprehension result partly from poor representations of a word's form and its meaning. Segalowitz and Hebert (1990) also point out that the speed of recalling meanings of L2 vocabulary, which L1 readers would carry out without much trouble, is slower than that of L1 vocabulary. For example, in reading an abstruse L2 text including some vocabulary of which most readers would not know the meaning, they could not conduct proper language processing, leading to their failure to comprehend the text (Berquist, 1997).

In order to solve the problem of the spontaneous activation of several cognitive resources within the capacity limitations, it is crucial to enhance the automatization of decoding skills, that is, to increase the speed of decoding target words. In fact, many studies have already recognized the importance of the automatization of word decoding both in the L1 and L2 reading process (Balass, Nelson & Perfetti, 2010; LaBerge and Samuels, 1974; Perfetti, 1985; Samuels, 2004). These studies insist that automatic word decoding will result in freeing WM and allowing readers to pay more attention to comprehension processes, leading to better text

comprehension. Instructional practices can play an important role in automatizing word decoding and thereby freeing more WM for comprehension (Perfetti, 1985; Samuels, 2004).

On the basis of this claim, our study argues for the importance of more effective devices as functions in CALL than those used traditionally, and then focuses on how a certain CALL function may enhance the automatization of word decoding skills, or quick recall of the meaning of target words in L2 reading.

2. Our study

The automatization of L2 word decoding to free WM can be conducted within a multimedia learning environment. This is because recent multimedia learning environments are based mainly on Cognitive Load Theory (Sweller, 1994), which is based on the idea that WM is limited in capacity, so it is necessary to design a learning environment in which WM may be used effectively (Moreno & Park, 2010). Based on this theory, optimum learning occurs when the cognitive load on WM is kept to a minimum in order to enable storage of the necessary information for longer periods (Sweller, 1988; Al-Shehri & Gitsaki, 2010). Multimedia learning can enhance the L2 learning environment because multimedia learning involves not only verbal gloss, but also other types of glosses, such as pictures, animation, sound, and so on. In fact, many studies have been conducted in order to verify the effectiveness of multimedia in L2 vocabulary learning (Al-Seghayer, 2001; Brett, 1998; Chanier & Selva, 1998; Chun & Plass, 1996; Lomicka, 1998; Mayer & Anderson, 1991; Pachler, 2001; Sato & Suzuki, 2010; Yoshii & Flaitz, 2002). However, these studies focus on incidental vocabulary learning with multimedia after reading activities, not on the effectiveness of multimedia in intentional vocabulary learning leading to successful reading. In other words, the goal of previous studies might have been longer retention of the vocabulary in the text, but our study regarded vocabulary learning as a facilitator in freeing WM, leading to successful comprehension of an L2 text with multimedia functions. However, which multimedia function will facilitate vocabulary learning in terms of automatization of word decoding?

Our study hypothesizes that L2 vocabulary learning practices with limited response time would be preferable. We assume that vocabulary learning practices with learners being conscious of their response time could accelerate word-decoding speed. Tan and Nicholson (1997) also demonstrate that “by putting pressure on attentional capacity in order to decode words, less attention will be available to process the meaning of the text” (*op. cit.*: 276).

Such a time-control function would be relatively easy to implement in a CALL environment, for example, in a website, application for personal computers, mobile phones, etc., whereas Tan and Nicholson (1997) examined the use of paper-based flashcards. In fact, there are many different types of software with a time-control function for CALL (e.g., Adachi & Suzuki, 2008). Although the function is used to shorten users’ answering time, it has not been validated from a theoretical perspective in terms of whether the function could actually enhance the learning of target words, hasten learners’ access to the meaning of words, and finally facilitate

text comprehension. Therefore, our study conducted statistical research on the basis of the following research questions:

1. Can those who learn L2 vocabulary with CALL material which requires them to find the correct meaning within a time limit recall the meanings of the learned vocabulary more promptly than those who learn the words without such a time limit?
2. Can CALL software with a time-control function contribute to greater retention of the L2 vocabulary than learning the words without such a time limit can?
3. Can CALL software with a time-control function enhance comprehension of the text with the L2 vocabulary learned, owing to the greater availability of WM for comprehension?

In order to verify our hypothesis, the following statistical research was conducted. In the next section, the details of our research are illustrated.

3. Research methods

3.1. Hypothesis

Our research was conducted in order to validate the following hypothesis: L2 vocabulary learning with time-control technology helps to automatize decoding skills, to free WM, and as a result to enhance text comprehension. In concrete terms, we verify whether time-control technology can assist learners to recall more vocabulary meanings and shorten recall response times, and furthermore whether shorter response time for vocabulary recollection can facilitate comprehension of the text, as a result of more WM being available for text comprehension.

3.2. Participants

The participants in our research were from the university at which the first author works. They are Japanese EFL learners, from freshmen to postgraduates, and belong to the Department of Agriculture or the Department of Technology; therefore, they do not have a specialization in English or a foreign language, but they are obliged to take some English classes (reading, writing, and communication) during the first two years.

3.3. Procedures

Our research method includes five steps: 1. a reading test to divide the participants equally into two groups; 2. a vocabulary pre-test; 3. a vocabulary learning exercise with different learning tools for each group; 4. a reading test on the texts containing the vocabulary they have learned; 5. a vocabulary post-test with time measurement.

The research was conducted in the office of the first author, where a quiet environment was maintained so that the participants could concentrate on their learning activities. At most five participants at a time performed the experimental tests in the office; thus, all their activities were carried out under the close supervision of the investigator.

3.3.1 Reading test. First, a reading test was conducted in order to examine the reading proficiency of the participants. The reading materials were taken from Part Seven of the reading section of the TOEIC® test, which is a popular evaluation test among Japanese university students. The participants were asked to answer ten questions after reading four texts within ten minutes. Then, they were divided into two groups according to their test results: a control group and an experimental group. This confirmed that their L2 reading abilities were mostly the same, which made the verification with different treatments feasible.

3.3.2 Vocabulary pre-test. Next, the vocabulary pre-test was conducted. The participants were asked to write the meanings of twenty-nine English words in Japanese, which we chose from the text that they would read in the subsequent reading test. The participants could answer the questions at their own pace, without any time limit. These words were selected in an arbitrary manner; however, we did attempt to choose somewhat difficult words for them on the basis of our teaching experience. The order of these words for the test was randomized. No feedback was given after the test; thus, the participants did not know the correct meanings of these words at that time.

3.3.3 Vocabulary learning exercise. Next, the participants in each group were given different learning materials with which to learn the twenty-nine words they encountered in the vocabulary pre-test.

Each participant in the *experimental group* was provided with an iPod, with an application for learning the installed vocabulary. This application had a function that could configure the time in which learners must provide the appropriate translations of the word meanings. The subjects were required to provide all the correct translations within ninety seconds (approximately 5 seconds per word), after which the application would be terminated automatically; Participants were able to see the correct translation immediately after providing their own translation, and they were allowed to restart the questions after the shutdown. The application randomized the order of the words which appeared on the screen.

Figure 1 presents an image of a sample question on an iPod screen. Below the target word, three alternative translations were shown, one of which was the correct answer, the other two being distractors. The users of the application were asked to touch the number that they believed was the correct answer.

Each participant in the *control group* was given a sheet of paper. On one side were the twenty-nine target words with three possible answers for each; on the other side, the correct answers were listed. In other words, the participants in the control group were given exactly the same amount of information as those in the experimental group: the target English words, three distractors per word, and the correct answers as feedback.

The participants in the control group learned the words by choosing the correct answer from the three alternatives in a similar manner to those in the experimental group. However, there was no regulation in terms of the time it took to look up correct answers at the back and the required time to find the correct answers: the participants in the control group learned the vocabulary words in their own individual ways.



Fig 1. iPod application for target vocabulary

3.3.4 Reading test. The participants in both groups learned the twenty-nine words within ten minutes. After finishing their learning, the participants were asked to read two texts on academic topics and answer the questions attached to each text within twenty minutes. The two texts consisted of 157 words and 352 words, respectively.

The reading texts were derived from the Society of Testing English Proficiency (STEP) test 1st grade and TOEFL, with less specialized topics for the participants. The scores of the readability in terms of Coleman-Liau Index (CLI) were 15.7 and 14.7, respectively, which implies that the readability of both texts was rather similar. However, considering the readability scores indicate that the texts are appropriate for second- or third-year undergraduate L1 students, they seem rather difficult for L2 students of a similar age in a Japanese university.

The questions in each text were classified into groups of several questions for multiple-choice comprehension (four alternatives each) and a translation of a sentence that included some of the words the participants had learned through the iPad application or the paper vocabulary list. During the sentence processing required by the translation

task, learners have to process both the form and meaning of the input (Juffs & Harrington 2011). Therefore, the increase of WM resulting from practice with time-control technology mightenable successful processing of form and meaning.

3.3.5 Vocabulary post-test. Finally, the vocabulary post-test was implemented. The test consisted of the same words and tasks as the pre-vocabulary test; however, the participants were asked to measure the time they used to answer each question, with stopwatches, and write down the time to two decimal places alongside the answer. They did not have to measure their answering time if they could not remember the correct meaning of the target word. The order of the post-test was randomized, but it was different from that of the pre-test.

4. Results

First, we illustrate the result of the first reading test in order to confirm the equivalence of the two groups into which the participants were divided. As a result of the test (see figure 2), the average scores of the control group and the experimental group were 5.7 (SD = 1.01) and 6.0 (SD = 0.82), respectively. These results were analyzed using t-test, and no statistical significance was found ($p = 0.32 > 0.05$). This implies that there was no difference in either group in terms of their English reading competence. The result also illustrated that, in accordance with Berquist (1997), who states that the scores of TOEIC® test reading section correlate with the command of L2 WM, so that if there is no statistical difference between the scores of each group it means that there is no difference between their command of L2 WM. This may indicate that the findings of our research result from the effectiveness of the learning materials we gave to the subjects, not from individual differences in their command of WM.

We analyzed the results from ten words (averse, concerted, contagion, eavesdrop, fatality, intercept, stray, zap, turn in, combative) in the post vocabulary test, that is, the percentage of correct answers and the response time. The ten words were selected from the pre-test in which none of the participants could identify the correct answers,

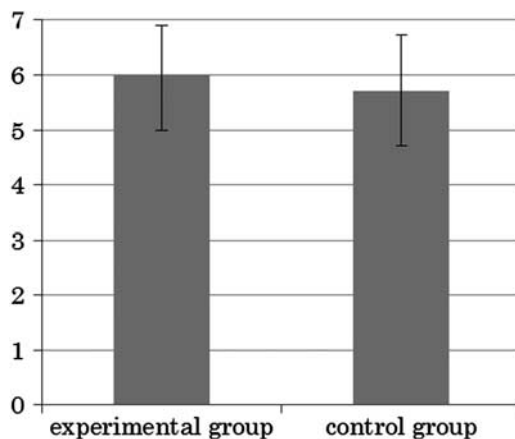


Fig 2. TOEIC® score results between the two groups

and they consisted of neither the Academic Words List (Coxhead, 2000), nor a pilot science-specific word list (Coxhead & Hirsh, 2007), both of which are used as an index of the vocabulary in academic texts. We analyzed words that none of the participants knew, which is vital for examining the effectiveness of the treatments given to each group. This is because correct data could not be obtained if the words were previously known before using the treatments.

4.1. The results of vocabulary recall

We will begin by analyzing the rate of correct answers. As the result of our analysis, seven out of ten words (averse, concerted, eavesdrop, intercept, stray, zap, turn in) were scored more by the experimental group than the control group (see Figure 3). Furthermore, with regard to the words for which the control groups received higher scores (contagion, fatality, combative), the differences between the groups were very small—0.01, 0.08, 0.07, respectively.

Next, the response times for the ten words were analyzed: Six words out of the ten words (concerted, contagion, eavesdrop, intercept, stray, combative) were answered quickly (see Figure 4). Moreover, the total average time of the response times showed that the experimental group answered the ten unknown words more quickly than the control group (6.66 > 7.01). As a result of the two-sided t-test (see Figure 5),

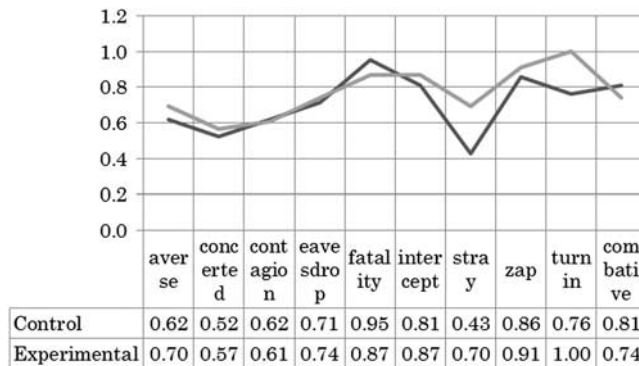


Fig 3. Average accuracy rates in the post-test (percentage)

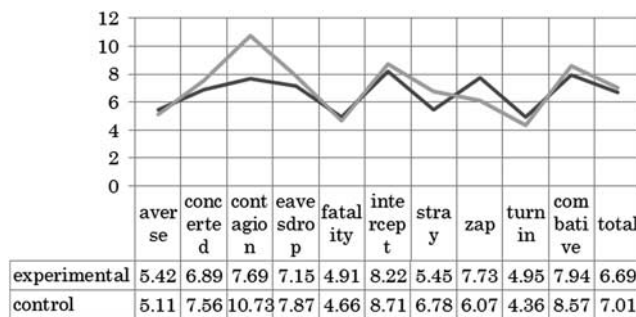


Fig 4. Average response times in the post-test (seconds)

two-sided	averse	concerted	contagion	eavesdrop	fatality	intercept	stray	zap	turn in	combative
t	0.48	0.73	2.53	1.33	0.40	0.48	1.34	0.87	0.85	0.50
df	27.0	22.0	28.0	30.0	37.0	35.0	9.9	32.7	35.7	20.4
p >.05	0.63	0.47	0.02*	0.19	0.69	0.63	0.21	0.39	0.40	0.62
method	Student	Student	Student	Student	Student	Student	Welch	Welch	Welch	Welch

Fig 5. Two-sided t-test results

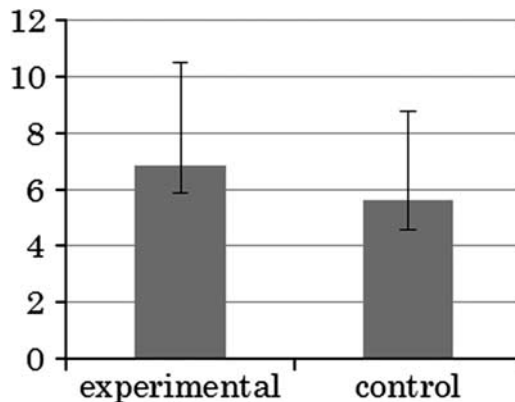


Fig 6. Total averages of reading test scores

one of the six words (contagion) demonstrated a significant difference ($p = 0.02 < 0.05$). On the other hand, the four words (averse, fatality, zap, turn in), which the control group answered quickly, did not show any significant difference in response time.

In order to sum up the two results mentioned above, higher scores and quicker responses were shown by the experimental group, although all results did not show statistically significant differences. However, one of the results from the experimental group illustrated a significant difference, whereas the results from the control group did not.

4.2. The results of reading comprehension tasks

Here, we discuss the results of the reading task containing the words the participants had learned. As mentioned above, the task required learners to read two English texts and answer several comprehension tests and a translation test on each text. We separately marked the translations on a scale of one to ten, according to the criteria we had discussed in advance. Then the average score was used as the examinee's score.

First, shown below are the total average scores of the reading test. As seen in Figure 6, the total average score from the experimental group was 6.87 (SD = 3.62), whereas that from the control group was 5.60 (SD = 3.16). The result of the two-sided t-test was that a marginal significance was found ($t(85) = 1.73, p = 0.09 > 0.05$).

Second, the results for each question on the reading tests is examined. The results of Reading Test 1 are illustrated in Figure 7. The average score for the comprehension test of the experimental group was 3.83 (SD = 1.44), while that of the control group was 3.14 (SD = 1.49). This implies that the experimental group obtained a higher score than

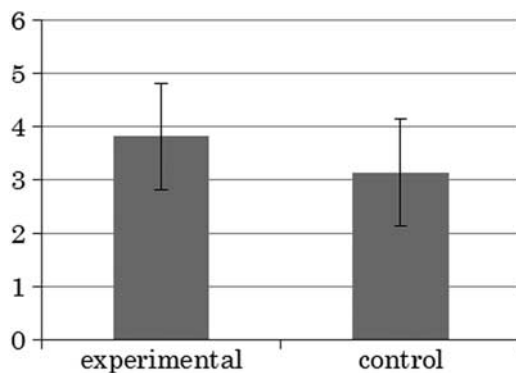


Fig 7. Average scores for Reading Test 1—comprehension test

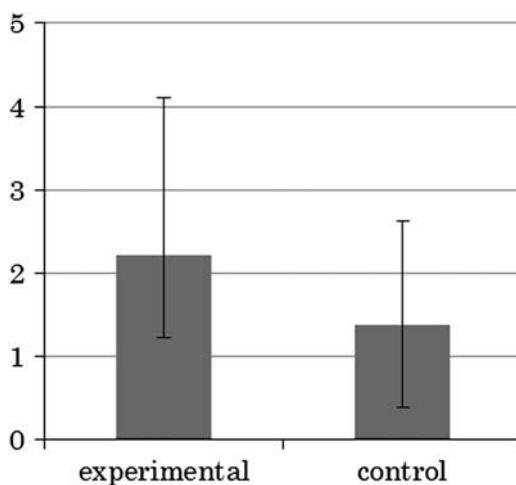


Fig 8. Average scores for Reading Test 1—translation test

the control group, although no statistically significant difference was shown ($t(42) = 1.55$, $p = 0.13 > 0.05$).

The average scores for the translation tests were 2.22 ($SD = 1.88$) for the experimental group and 1.38 ($SD = 1.24$) for the control group (see Figure 8), which represents a marginal significance found as a result of our two-sided t-test ($t(42) = 1.72$, $p = 0.09 > 0.05$).

With regard to the total average scores of Reading Test 1, the experimental group scored 6.04 ($SD = 2.88$), while the control group scored 4.52 ($SD = 2.25$), which shows that the experimental group obtained a higher score (see Figure 9). As a result of our two-sided t-test, there was a marginal significance between the groups ($t(42) = 1.94$, $p = 0.06 > 0.05$).

Next, the results of Reading Test 2 are illustrated. As seen in Figure 10, the average score for the comprehension test of the experimental group was 0.96 ($SD = 0.88$), whereas that of the control group was 0.33 ($SD = 0.58$). As a result of the two-sided t-test, a statistically significant difference was found between the

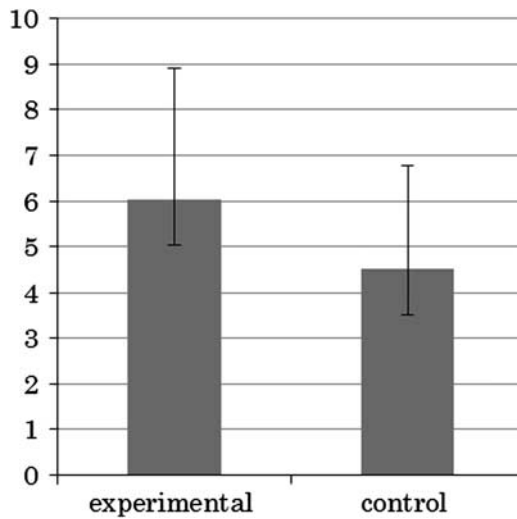


Fig 9. Average scores of Reading Test 1

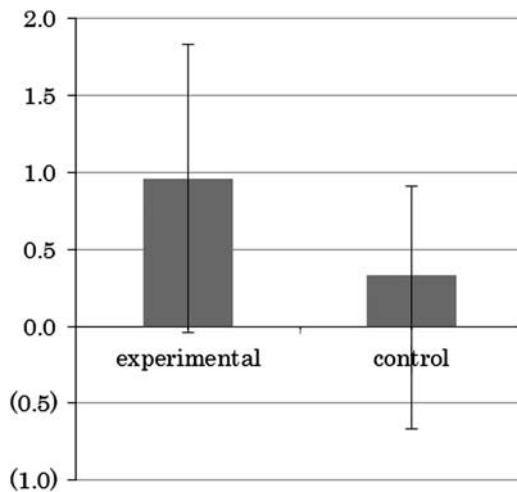


Fig 10. Average scores for Reading Test 2—comprehension test

groups ($t(42) = 2.75, p = 0.01 < 0.05$). The result showed that the experimental group's score was significantly higher than that of the control group.

With regard to the translation test in Reading Test 2 (see Figure 11), the average score of the experimental group was 6.74 ($SD = 3.65$); on the other hand, the average score of the control group was 6.21 ($SD = 3.40$). This demonstrated that the experimental group had a higher score, as with the other reading tests. However, according to our two-sided *t*-test, there was no statistically significant difference between the groups ($t(42) = 0.49, p = 0.62 > 0.05$).

Finally, with regard to the total average scores of Reading Test 2, the experimental group scored 7.70 ($SD = 4.13$), while the control group obtained 6.55 ($SD = 3.63$),

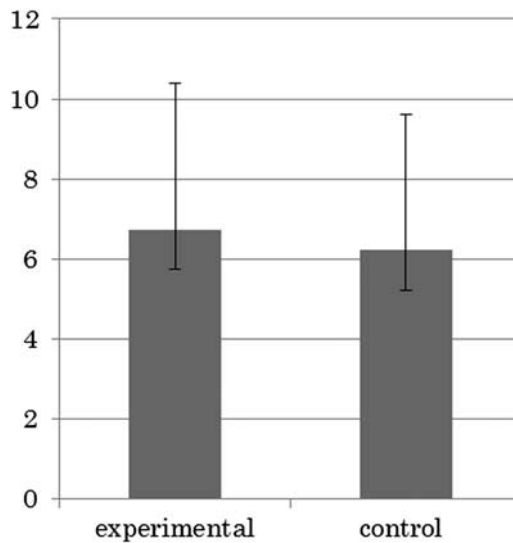


Fig 11. Average scores of Reading Test 2—translation test

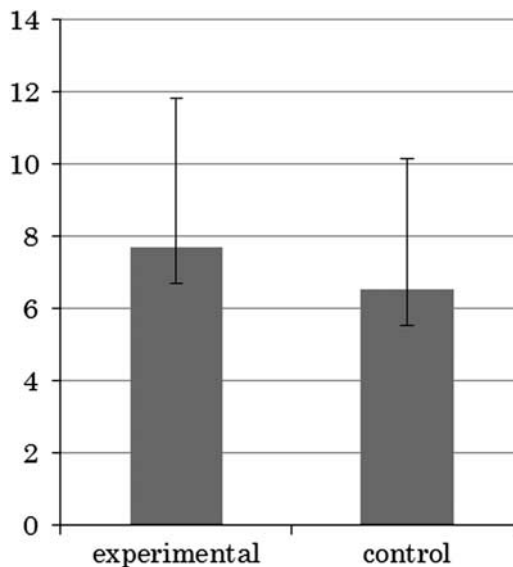


Fig 12. Average scores of Reading Test 2

which showed a higher score for the experimental group (see Figure 12). However, according to the result of our two-sided t-test, no significant difference was shown between the groups ($t(42) = 0.98$, $p = 0.33 > 0.05$).

To sum up, we conducted four reading tests consisting of multiple comprehension tests and a translation test from two English texts, respectively. The experimental group not only obtained higher scores in all four tests, but also demonstrated two marginally significant and one statistically significant difference out of the four tests.

5. Findings

In this section, we examine the three hypotheses proposed in this study. Our first hypothesis that the language learning device with a time-control function may facilitate L2 vocabulary learning, leading to quicker recall of the meanings of 29 target words, is supported by the findings that the experimental group obtained not only higher accuracy rates but also higher average scores for reaction times. In addition, the experimental group showed statistically significant differences in some of the average scores and response times. Considering these results, we can say our first hypothesis might be generally supported by the results of our experimental research. These findings support our claim that time-control technology can facilitate the automatization of word decoding skills and thereby increase the amount of WM available for comprehension.

Our second hypothesis is that L2 vocabulary learning with time control technology can bring about greater retention of the meaning of the target words. This hypothesis is also supported by higher scores shown by the experimental group. This result might also arise from the automatization of decoding skills resulting in the availability of more WM for recalling the meanings of words.

Our third hypothesis is whether foreshortening the response time to recall word meanings could facilitate better comprehension of the text containing the words. The findings indicate that the experimental group got the higher average scores in all the reading tests, one of which shows statistically significant differences, although the other average scores of the experimental group did not show statistically significant differences. These findings support our hypothesis that the WM freed by the time control technology can be used for comprehension, leading to better understanding of the text.

6. Conclusion and future suggestions

Our study began with the premise that L2 reading activities require more limited-capacity language processing than L1 reading activities, particularly bottom-up decoding skills, such as recall of the meaning of vocabulary within the text. On the basis of the idea that a multimedia environment helps readers facilitate the automatization of decoding skills and free WM, thereby resulting in successful comprehension of the text, we hypothesized that L2 vocabulary learning with a time-control technology would minimize readers' response times; therefore, more WM can be used for comprehension, thereby leading to better understanding of L2 texts. According to our statistical research, L2 readers who used an iPod application with the time-control technology, not only recalled more words but also recalled meanings more quickly than those who used a paper list on which the target vocabulary was written. We can therefore claim that the time-control technology can play a pivotal role in facilitating the automatization of L2 word decoding skills to free WM.

In addition, as a result of reading comprehension tests, those who used the time-control technology more often achieved higher scores on all questions, one of which was statistically significant. Although not all results are statistically significant, we may comprehensively conclude that vocabulary learning with a time-control

function facilitates not only longer retention of the target vocabulary words and quicker recall of their meanings, but also better comprehension of the texts. Taking the results into consideration, we can verify that the WM saved by the automatization is successfully used for text comprehension.

Although our test results support our three hypotheses overall, it is also true that modifications are needed for our research. For example, the results of this study could not lead to a strong assertion that the time-control function provided by CALL devices can facilitate the response time of recall in addition to text comprehension in terms of statistically significant differences. This might be partly because of the relatively small number of participants in our study, thereby preventing a strong conclusion.

Furthermore, the number of the target words we analyzed was rather small, despite the fact that we attempted to use difficult words that none of the participants would answer correctly in the pre-vocabulary test. In fact, the accuracy rate of the participants was better than we had expected; therefore, we could not analyze more than ten words.

Finally, there is some possibility that there was an uneven display of the words the learners were supposed to study using the CALL software, whereas the vocabulary list that the control group used could display any vocabulary word evenly on the sheet. Therefore, the learners could study every vocabulary word with the same effort. This uneven display of vocabulary might have occurred when those in the experimental group who used the software in the iPod touch could not study the vocabulary evenly if they could not finish answering all the vocabulary within the time-limit and then had to restart from scratch. The time-limit function was set to display the vocabulary at random; however, in this study, the learning time with the devices was so limited that the learners of the experimental group might not have studied all the target vocabulary evenly, as compared with the wordlist for the control group. This might be one reason the experimental group could not obtain many statistically significant differences in their accuracy rates and response times. Nevertheless, higher accuracy rates and average scores obtained by the experimental group could demonstrate the potential of the time-limit function of the CALL software to facilitate quicker recall. More deliberate modifications are needed for future research.

In this study, we shed light on text comprehension in reading; however, in order to further evaluate the effectiveness of the time-limit function of CALL, future research might be approached from another standpoint, such as from the perspective of listening.

References

- Adachi, H. and Suzuki, A. (2008) *Learning English Vocabulary: Quick and Fun!* Tokyo: Chukei Shuppan.
- Adams, M. J. and Collins, A. (1979) A schema-theoretic view of reading. In: Freedle, R.O. (ed.), *New directions in discourse processing*, 1–22. Norwood, New Jersey: Ablex Publishing Corporation.
- Al-Seghayer, K. (2001) The effect of multimedia annotation modes on L2 vocabulary acquisition: a comparative study. *Language Learning and Technology*, 5(1): 202–232.
- Al-Shehri, S. and Gitsaki, C. (2010) Online reading: a preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. *ReCALL*, 22(3): 356–375.
- Baddeley, A. D. (1986) *Working Memory*. Oxford: Clarendon.

- Balass, M., Nelson, J. R. and Perfetti, C. A. (2010) Word learning: an ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, **35**: 126–140.
- Berquist, B. (1997) Individual differences in working memory span and L2 proficiency: Capacity or processing efficiency? In: Sorace, A., Heycock, C. and Shillcock, R. (eds.), *Proceedings of the GALA '97 Conference on Language Acquisition*. Edinburgh, UK: Human Communication Research Center, University of Edinburgh, 468–473.
- Brett, P. (1998) Using Multimedia: A Descriptive Investigation of Incidental Language Learning. *Computer Assisted Language Learning*, **11**(2): 179–200.
- Cain, K. and Oakhill, J. (2007) Reading comprehension difficulties: Correlates, causes and consequences. In: Cain, K. and Oakhill, J. (eds.), *Children's comprehension problems in oral and written language: A cognitive perspective*. New York: The Guilford Press, 41–75.
- Carpenter, P. A. and Just, M. A. (1989) The role of working memory in language comprehension. In: Klahr, D. and Ktovsky, K. (eds.), *Complex Information Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 31–68.
- Chanier, T. and Selva, T. (1998) The ALEXIA system: the use of visual representations to enhance vocabulary learning. *Computer Assisted Language Learning*, **11**(5): 489–521.
- Chun, D. L. and Plass, J. L. (1996) Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, **80**(2): 183–198.
- Coxhead, A. (2000) A new academic word list. *TESOL quarterly*, **34**(2): 213–238.
- Coxhead, A. and Hirsh, D. (2007) A pilot science word list for EAP. *Revue Française de linguistique appliquée*, **xii**(2): 65–78.
- Gough, P. B. and Tumber, W. E. (1986) Decoding, reading and reading disability. *Remedial and Special Education*, **7**: 6–10.
- Harrington, M. and Sawyer, M. (1992) L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, **14**: 25–38.
- Juffs, A. and Harrington, M. (2011) Aspects of Working Memory in L2 learning. *Language Teaching*, **44**(2): 137–166.
- LaBerge, D. and Samuels, S. J. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology*, **6**: 293–323.
- Lomicka, L. L. (1998) “To Gloss or Not To Gloss”: An Investigation of Reading Comprehension Online. *Language Learning & Technology*, **1**: 41–50.
- Mayer, R. E. and Anderson, R. B. (1991) Animations need narrations: An experimental test of a dual-code hypothesis. *Journal of Educational Psychology*, **83**: 484–490.
- Moreno, R. and Park, B. (2010) Cognitive Load Theory: Historical Development and Relation to Other Theories. In: Plass, J., Moreno, R. and Brünken, R. (eds.), *Cognitive Load Theory*. New York: Cambridge University Press, 9–28.
- Nation, K. and Angell, P. (2006) Learning to read and learning to comprehend. *London Review of Education*, **4**(1): 77–87.
- Oakhill, J. V., Cain, K. and Yuill, N. M. (1998) Individual differences in children's comprehension skill: Towards an integrated model. In: Hulme, C. and Joshi, R. M. (eds.), *Reading and spelling: Development and disorder*. Mahwah, NJ: Lawrence Erlbaum Associates, 343–367.
- Pachler, N. (2001) Electronic reference tools for foreign language learners, Teachers and users: Offline vocabulary look-up programs. *Language Learning Journal*, **24**: 24–29.
- Perfetti, C. A. (1985) *Reading ability*. New York: Oxford University Press.
- Perfetti, C. A. and Lesgold, A. M. (1977) Discourse comprehension and source of individual differences. In: Just, M. A. and Carpenter, P. A. (eds.), *Cognitive Processes in Comprehension*. Hillsdale, NJ: Erlbaum, 141–183.

- Samuels, S. J. (2004) Toward a theory of automatic information processing in reading, revisited. In: Ruddell, R. B. and Unrau, N. J. (eds.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association, 1127–1148.
- Sato, T. and Suzuki, A. (2010) Verifying multimedia gloss: Image schemata and polysemous vocabulary in English. *Proceeding of EUROCALL 2009 Conference*, Valencia, Spain: McMillan ELT, 285–294.
- Segalowitz, N. and Hebert, M. (1990) Phonological coding in the first and second language reading of skilled bilinguals. *Language Learning*, **40**: 503–538.
- Stuart, M. (2006) Learning to read: developing processes for recognizing, understanding and pronouncing written words. *London Review of Education*, **4**(1): 19–29.
- Sweller, J. (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Science*, **12**: 257–285.
- Sweller, J. (1994) Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, **4**: 295–312.
- Tan, A. and Nicholson, T. (1997) Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*, **59**: 276–288.
- Yoshii, M. and Fraitz, J. (2002) Second Language Incidental Vocabulary Retention: The Effect of Text and Picture Annotation Types. *CALICO Journal*, **20**(1): 33–58.