THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE: Serrano, R., & Huang, H. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly, 52(4)*, 971-994.

Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text?

Abstract

Repeated reading, which involves the reading of short passages several times, has been demonstrated to be beneficial for second language fluency (Chang & Millett, 2013) and vocabulary acquisition (Liu & Todd, 2014). Despite the increasing interest in repeated reading, no study has addressed the effects of time distribution—how different encounters with the same text should be spaced for repeated reading to have the strongest impact on second language learning, specifically on vocabulary acquisition, the focus of the present study. This study includes two groups of 16-year-old EFL learners in Taiwan (n = 71). One group carried out assisted repeated reading (i.e., with audio support) once every day for five consecutive days (intensive distribution); the other read the same text once every week for five consecutive weeks (spaced distribution). Our results revealed that intensive practice led to more immediate vocabulary gains but spaced practice led to greater long-term retention.

Introduction

Incidental vocabulary learning through reading

Considering the vast number of words that learners need to acquire, it seems quite

unrealistic to expect them to be learned explicitly in the classroom. Although explicit techniques of vocabulary learning have been shown to be more effective in promoting word knowledge than more implicit methods (Laufer, 2006), several studies have also shown that incidental vocabulary learning through reading can also occur (Pellicer-Sánchez & Schmitt, 2010). It must be pointed out that the term "incidental learning" has been operationalized in different ways. According to Hulstijn (2013), incidental learning refers to "the acquisition of a word or expression without the conscious intention to commit the element to memory" (p. 2632). Our study was designed to create conditions for incidental vocabulary learning in the sense described by Peters et al. (2009): "*Incidental* entails that participants are not forewarned of an upcoming test, whereas the label *intentional* refers to the fact that participants are explicitly told that an upcoming test will follow" (p. 116). We do not have any concurrent data to support the claim that the learning that took place in our study was indeed incidental, and that is why we are adhering to this purely methodological operationalization of the term (see more details about the procedure in the methodology section). It must be emphasized, however, that participants do not always behave as expected concerning the experimental condition they have been assigned to, and some might engage in intentional learning even in supposedly incidental learning conditions (DeKeyser, 1995; Robinson, 2002; 2005).

The degree of incidental vocabulary learning from reading that has been reported in previous research varies from study to study. There is some evidence, however, that seems to suggest that the potential of reading to promote vocabulary learning can be increased when it also includes audio support (assisted reading), or when learners have

opportunities for repeated reading practice.

Webb and Chang (2015) examined the effect of extensive assisted reading on incidental vocabulary learning with Taiwanese secondary school students who read 10 level-one graded readers over 13 weeks. The students were tested on 100 target words selected quasi-randomly from the readers in a vocabulary matching test. The study also included a control group that followed their regular form-focused EFL instruction. Their results showed that incidental vocabulary gains through reading and listening to multiple texts were high, with an absolute gain of 19.72 words, which represents 39.09% of the words unknown on the pretest. In contrast, the control group experienced an absolute gain 4.43 words.

The same authors also analyzed the effect of audio support on repeated reading and compared the effect of assisted repeated reading and unassisted repeated reading on incidental vocabulary acquisition in a real classroom (Webb & Chang, 2012). The results of two vocabulary tests revealed that beginner readers could acquire new receptive vocabulary knowledge incidentally through both methods but assisted repeated reading led to significantly greater gains than unassisted repeated reading. According to the authors, the prosodic features of assisted repeated reading help students segment the information in more meaningful chunks, which benefits overall comprehension, and greater overall comprehension could in turn facilitate guessing of unknown words. Additionally, the connection between written and oral forms of words in assisted repeated reading provides greater memory links, which might also encourage vocabulary learning.

Liu and Todd (2014) further examined the effects of assisted repeated reading on

incidental vocabulary learning. Eighty high-intermediate Taiwanese learners of Japanese were recruited in a university to participate in one assisted repeated reading session. In the experiment, they either had to read one single passage in Japanese of around 1,200 words seven times, or seven different passages of equal length, which included the same target words. The results of the study showed that, on average, the students learned 10.47 words, which is equivalent to 29% of the new words (n = 36). The difference between pre- and post-test was statistically significant, and the students that read seven different passages learned more than those that read the same passage seven times.

An additional variable that should be taken into account when analyzing the effect of repeated exposures to target words in repeated reading is the distribution of those exposures. On the one hand, it is relevant to explore whether it is better to provide repeated exposure to the same words in only one session or in different sessions. When several sessions are considered, it is important to analyze whether different intersession intervals could also result in differential learning gains.

Time distribution and learning

In cognitive psychology, there is a long tradition of research on the effect of time distribution on learning with one of the most robust findings being the spacing effect or distributed practice effect (see reviews in Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Rogers, 2017; Rohrer, 2015; Serrano, 2012). As the name indicates, the spacing effect suggests that learning is enhanced when the study time is distributed in spaced

sequences rather than in massed sequences, which generally involve one session where the material to be learned is repeated within short time intervals. Most of the reported findings on the spacing effect are based on laboratory studies, where participants were exposed to the material to be learned (typically words) through a computer screen at different time intervals, often in one session.

Contrary to empirical evidence in favor of distributed practice, exploratory studies investigating different time distributions in second language acquisition seem to generate different results (Collins, Halter, & Lightbown, 1999; Serrano & Muñoz, 2007; Collins & White, 2011; Serrano, 2011, 2012). In general, intensive L2 courses (which concentrate many hours of instruction in a limited time, including long sessions with short intersession intervals) have been found to be more beneficial for general L2 learning (e.g., reading skills, listening skills) than courses that space the same hours throughout longer periods (thus offering shorter sessions that are more widely spaced). The results of these studies, though, are not conclusive, as in some cases factors other than time distribution could have been at play (Collins et al., 1999; Collins & White, 2011), or in other cases the advantages of intensive instruction do not apply to all the L2 aspects examined or to all proficiency levels (Serrano, 2011; Serrano, Stengers & Housen, 2015).

It is interesting, though, that these results run counter to the findings in the cognitive psychology literature. Researchers have provided several possibilities for these seemingly contradictory findings (Rogers, 2017; Author2, 2012), and one may be the operationalization of massed versus spaced. Rogers (2017) indicates that in the SLA literature the learning analyzed occurs in multiple training sessions which differ

in terms of the lag between learning sessions. In cognitive psychology experiments, on the other hand, learning often takes place in one single session. As Rohrer (2015) suggests, the relevant comparison in educational settings is not between *massing* in one single session versus spacing the learning in different sessions, as classroom learning typically involves more than one session. In the SLA literature, the interest is not so much in the spacing effect, but rather, in the *lag effect* (Rogers, 2017), or how much time there should be between different learning sessions.

Another variable that needs to be considered when analyzing the results from the time distribution literature is when testing takes place. The exploratory SLA studies mentioned above examine only short-term learning and do not investigate long-term retention through delayed post-tests. There are, in fact, some studies in the cognitive psychology literature that have failed to confirm the spacing effect at short retention intervals (Bahrick & Hall, 2005; Rohrer & Taylor, 2006). Toppino and Gerbier (2014) suggest that longer spacing between sessions tend to be especially beneficial for long-term learning; however, when the retention interval is too long, forgetting might take place (Cepeda et al, 2006).

The activation-based model of the spacing effect proposed by Pavlik and Anderson (2005) tries to account for the lack of spacing effects at short retention intervals. According to the authors, each time an item is practiced, its mental representation is strengthened, but these increments of strength decay as a power function of time. The rate of decay will be determined by the degree of activation at the time of the presentation. In massed sequences, activation of the first presentation of an item (P1) at the time of the second presentation (P2) is high, thus preventing encoding

of P1 into long-term memory. When activation of P1 is weak at P2, long-term memory encoding occurs. Consequently, the higher the degree of activation at the time of the presentation (as in massed sequences), the faster the decay rate will be, or, in other words, the shorter time that presentation will be retained. Therefore, a sharp learning increase due to high activation in massed sequences is expected to lead to a faster knowledge decline.

Rohrer and Pashler (2007) explored the interaction between the learning and the retention schedules and suggest that, depending on when testing takes place (the retention interval, or interval between the end of the treatment and the test), the learning outcomes of different types of spacing (intersession intervals) would be better or worse. The authors claim that the intersession interval should be between 10 and 30% of the retention interval. For instance, if testing takes place within a retention interval (RI) of 10 days (i.e., 10 days after the treatment), 1-day intersession interval (ISI) would be 10% of the RI and thus a good spacing to analyze learning.

With a more concrete example, if a teacher wants to teach the present perfect in three sessions which will be on three consecutive days (i.e., 1-day ISI), a test 10 days after the treatment (10-day RI) would be appropriate to measure learning, according to Rohrer and Pashler (2007). However, if that same teacher decides to include 7 days between each learning session, and have one session per week during three weeks (7-day ISI), having the test within the same retention interval (10 days after the treatment) would not probably capture well the learning outcomes of that program. The ISI (7 days) would be 70% of the RI (10 days), and thus not optimal according to the authors (see Figure 1). In this case, a test that takes place 30 days after the treatment

(30-day RI) would be better, as the ISI (7 days) would be 23% of the RI (30 days), and within the range proposed by Rohrer and Pashler (2007).

[Figure 1]

Apart from the previously mentioned SLA studies, which mostly examine general L2 skills, more controlled studies have also been performed in order to further investigate the role of time distribution in L2 learning, and they have taken into account the claims made by Rohrer and Pashler concerning ISI and RI (see Table 1 for a summary). Bird (2010) explored the effects of distributed (14-day ISI) and concentrated (3-day ISI) practice on the retention of English syntax by Malay-speaking university students at two intervals: 7-day RI and 60-day RI. The results of the grammaticality judgment tests revealed that there were no significant differences between the two groups in the 7-day post-test. However, at a 7-day RI, the ISI was not optimal for either the distributed (ISI/RI ratio = 200%) or the concentrated (ISI/RI ratio = 42.85%) program, following Rohrer and Pashler (2007). On the other hand, at a 60-day RI, the 14-day ISI group significantly outperformed the 3-day ISI group (with a large estimated effect size). In this case, the RI was optimal for the distributed program (ISI/RI ratio = 23%), but not for the concentrated (ISI/RI ratio = 5%).

Rogers (2015) obtained similar results in the case of incidental English L2 grammar learning assessed through a grammaticality judgment test. No differences were found in the immediate post-test between the concentrated (ISI = 2.25 days) and the spaced (ISI = 7 days) groups. However, there was a significant difference between the two groups for long-term retention (RI = 42 days) (with a medium effect size), in

favor of the spaced group. As was the case in Bird's study, the RI for the delayed post-test was only optimal for the spaced group.

Suzuki and DeKeyser (2017a) examined the acquisition of a Japanese morphological structure by beginner-level US university students through two learning sessions in two conditions: 1-day ISI (concentrated group) and 7-day ISI (distributed group). The authors then analyzed retention 7 days and 28 days after the end of the treatment, in terms of accuracy and speed in oral production of the form. The results indicate that the concentrated group outperformed the distributed group in terms of accuracy at both post-tests (although the differences were not statistically significant). Considering speed, there was one significant interaction between time and group in favor of the concentrated group, especially in the delayed post-test (28-day RI), with a medium effect size. The researchers argued that the fact that their study probably taps more procedural than declarative knowledge could explain the differences in results between their study and Bird's.

In a replication study, Suzuki (2017a) examined the learning of morphology of a novel miniature language under shorter (3.3-day ISI) and longer (7-day ISI) spacing conditions. His results provide further support for the higher learning outcomes of the participants in the shorter ISI condition, although in this case the significant advantages were found for accuracy and not for speed. Suzuki (2017b), in a reanalysis of the previous study, provided further evidence for the advantages of the shorter ISI group, this time in terms of automatization of morphological rules; although, like in previous cases, the differences between groups were not statistically significant in all the comparisons.

[Table 1]

The current study

It appears that time distribution has a relevant role; however, we still need more studies in order to learn more about the optimal intersession intervals for the acquisition of different L2 skills, as the results of previous studies seem to present different (even contradictory) evidence. In the area of assisted repeated reading, teachers might ask is how the readings of the same text should be spaced in order to maximize the learning potential of assisted repeated reading.

This study examines how different ISIs affect short-term learning and retention of vocabulary through assisted repeated reading. Considering the potential effect of the RIs on the results, we have decided against one single scheduling for the delayed post-test (as in Bird, 2010 or in Rogers, 2015), and instead to test retention in conditions that fall between what Rohrer and Pashler (2007) would consider optimal ranges for the two groups under analysis. These ranges provide a mathematical relation between ISI and RI, which still needs to be confirmed by other studies. However, they are a starting point for researchers to consider for their designs when planning testing. Many authors have observed that the benefits of concentrated learning tend to show at short retention intervals while learning in spaced sequences appears to be more advantageous for long-term learning (see Cepeda et al., 2006). Rohrer and Pashler's study tries to define a bit clearer what "ideally short" and "ideally long" might be. It must be emphasized, though, that the purpose of this study is not to look into optimal

ISI/RI ratios (for that we would need to include different ones), but to examine learning and retention in the fairest possible way. Including the same lag between post-test and delayed post-test for the two groups may involve choosing a RI that is too short for the benefits of distributed learning to show, or too long for concentrated learning to have a positive effect.

The research questions that guide this study are:

1. Can assisted repeated reading effectively promote incidental vocabulary learning and retention?
2. Does the time distribution of reading sessions of assisted repeated reading (intensive vs. spaced) have an effect on vocabulary learning and retention?

Method

Participants

The participants consisted of 80 EFL students from two intact classes in a high school in Taiwan (grade 10, ages 15-16). One class was randomly assigned to an intensive treatment, and the other a spaced treatment. A total of 71 students ended up participating in all session, 37 in the intensive and 34 in the spaced group.

Seven months prior to the study, the students were required to take an entrance exam on general achievement for all academic subjects in order to be placed in different high schools. Hence, the students in the same school can be considered to have similar academic aptitude and abilities. When this study started, the students were in the second semester of grade 10, and they had been learning English for three years. All participants had had six 50-minute English classes each week (for a total of five

hours a week) in grade 10. During the first semester, they took three written exams for English class and the scores were used to examine whether the students in the two groups were comparable in terms of English proficiency. There were no significant differences between the two classes on the average score of these tests (M = 61.11/100, SD = 15.64 for the intensive *vs*. M = 66.82/100, SD = 12.27 for the spaced): $t(78) = -1.816$, $p = .073$, with a small effect size, ($d = -.41$), so the groups can be considered comparable. The classes were taught by two different teachers, so, in order to minimize a teacher effect, they were asked to follow the same procedure and read out the same instructions while leading the assisted repeated reading sessions.

Instruments

Treatment instrument: The reading passage. As materials selected for assisted repeated reading are often between 300 and 400 words, a short passage was chosen for this study. The text was from the English textbook (Chou et al., 2010) used in the high school where the study took place, which has six volumes. The students were using volume 2 at the time of data collection, and they would use volume 3 the following semester (after data collection was over). The chosen passage was from volume 3 and was shortened and simplified by the second author to match the students' current level of English. The two teachers who taught the experimental classes deemed it suitable in terms of length, vocabulary, and difficulty for the participating students.

The length of the modified passage was 419 words. The topic was about Kyoto, introducing some cultural features of Japan. Due to Taiwan's proximity to Japan and the fact that Japanese is a required course in this high school, students had a certain

degree of knowledge of Japanese culture. In a questionnaire distributed to some senior students of the same school prior to a piloting period, this topic was deemed interesting to them.

The selected passage targeted 33 words. In order to make sure that the modified text and the originally selected target words were appropriate for the participants of this study, a pilot test was carried out with another grade 10 class in the same high school (n = 40), which included nine comprehension questions and a vocabulary test of the 33 words. The results of the pilot test indicated that the students could understand the general meaning of the text and could successfully answer the comprehension questions. The students knew fewer than 15 out of the 33 words indicating there was room for learning. One word that all the students knew was discarded and four new words, which the teachers expected to be unknown by many students, were added, making a total of 36 target words.

Regarding lexical coverage, for a grade 10 student who did not know any of the 36 target words in the current study, the coverage of the tokens was approximately 86%. Yet the coverage was higher for most students since, as previously explained, they were expected to know a certain number of target words. The coverage was calculated considering the total number of tokens included in the passage (419) minus 58 (word tokens representing the 36 target word types). Though this vocabulary coverage was well under the 98% suggested by Nation (2001) for reading comprehension, the reading passage was still considered suitable for the assisted repeated reading sessions because the students would be able to use a glossary of the target words. The glossary provided Chinese meanings for each target word as well as

three phrases appearing in the passage, which were considered potentially difficult for students. An audio recording of the reading was performed by a native English speaker at a slow speed, and lasted for approximately four minutes.

Testing Instruments

Vocabulary Levels Test. Following Webb and Chang (2012), the unpublished bilingual English-Chinese version of the Vocabulary Levels Test (VLT) designed by Anna Chang (drawing on previous VLTs designed by Paul Nation and by Schmitt, Schmitt & Clapham, 2001) was used as a measure of English proficiency. After piloting with three grade 10 students who did not participate in the study, the test up to 3,000-word level was given to the participants. Each 1,000-word level test consisted of 30 multiple-choice questions divided into six blocks. In each block, students had to match three Chinese words from six possible English equivalents. The 3,000-word VLT contained a total of 90 multiple-choice questions and took approximately 20 minutes to complete.

Each correct answer in the VLT was given a score of one with the highest possible score of 30 in each 1,000-word level. Following Webb and Chang (2015) and Chang and Millet (2013), students' scores on each 1,000-word level were multiplied by 33.3 and added up to get a final score, as an approximate measure of students' vocabulary size considering the most frequent 3,000 words.

Bilingual vocabulary matching test (BVMT). To assess vocabulary learning, a vocabulary matching test based on the one used by Webb and Chang (2015) was adopted in this study. The same bilingual matching test was used in the pre-test,

post-test, and delayed post-test. The test of 36 target words consisted of three blocks based with word categories (two blocks of nouns, and one block of adjectives and verbs) and their Chinese translations with two distractors provided in the opposite column in random order. The students had to choose the correct Chinese translation for each word and fill it in the blank. An example of one of the blocks is shown below:

_____	atmosphere	a. 護身符；符咒 (charm)
_____	district	b. 簡單 (simplicity)
_____	arrangement	c. 樂器 (instrument)
_____	shrine	d. 禪；禪宗 (Zen)
_____	present	e. 佛寺 (temple)
_____	packet	f. 氣氛 (atmosphere)
_____	charm	g. 茶道 (tea ceremony)
_____	instrument	h. 觀看者 (viewer)
_____	Zen	i. 神社 (shrine)
_____	viewer	j. 地區 (district)
_____	wood	k. 現在；現今 (present)
_____	simplicity	l. 小包 (packet)
		m. 木頭；木材 (wood)
		n. 排列 (arrangement)

The reliability of the instrument using Cronbach alpha suggests that the BVMT is reliable (0.795). As can be inferred from the results of the pilot study explained above, the final vocabulary test included words that some students would know. However, we agree with Webb and Chang (2015) that it is beneficial to include some known words in vocabulary tests as they may have the benefit of "encouraging engagement" (p.7). The time needed to complete the test was approximately 10 to 15 minutes.

Of the 36 target words in the test, each correct answer was given a score of one,

incorrect answers received 0 points. The highest possible score for this test was thus 36.

Design and Procedure

The study followed a pre-test, treatment, post-test, and delayed post-test design. The two groups, intensive and spaced, followed the same procedure for both the assisted repeated reading treatment and the vocabulary testing (see Figure 2).

[Figure 2]

Treatment procedure. The treatment included five sessions of assisted repeated reading followed by comprehension activities. The ISIs examined by Suzuki and DeKeyser (2017a) (1-day and 7-day) were adopted in the current study, as this time arrangement could be easily implemented considering the school calendar. The intensive group had one session a day (integrated within the English class) for five consecutive days over a week (1-day ISI) and the spaced group had one session a week over five consecutive weeks (7-day ISI). Previous research on RR often uses five repetitions in their design and even suggests that five repetitions help comprehension and vocabulary acquisition (Chang & Millett, 2013; Gorsuch, 2015). Similarly, the SLA research on distributed learning that has shown advantages for distributed conditions has also used five repetitions: Bird (2010), Rogers (2015) (vs. Suzuki & DeKeyser, 2017a). The glossary was given to students only after they had finished reading and listening to the passage, so that they would not focus on those words while reading the text.

In each session, the teacher distributed the reading passage first and the students were instructed to pay attention to the reading, while listening to it at the same time, and were encouraged to focus on understanding the content without worrying about unknown vocabulary. After reading while listening, the students were given the reading comprehension activities together with the glossary to refer to when necessary. The activities aimed to check if students understood the main ideas. In each comprehension set (there was a different set of questions for each assisted repeated reading session), there were five to six questions concerning general understanding of the reading and specific details. The use of the comprehension sheets was to ensure students focused on meaning when they read and listened to the text, but students' performance in this activity was not analyzed in this study. All the comprehension questions (including true and false questions, multiple choice, and fill-in-the blank) stimulated students to process the new vocabulary. The students needed to revisit each target word at least once in order to complete the reading comprehension activities. Each session took 15 to 20 minutes to complete. The last session also included an immediate post-test.

Testing procedure. Two days prior to the treatment, the two groups completed the BVMT and the participants had no English instruction during the two-day interval between the pre-test and the first assisted repeated reading session. At the beginning of the first treatment session, the VLT was administered. In the last session, after the reading and comprehension questions, all the students completed the same BVMT as the post-test. Students were given unlimited time to complete the tests. As in some previous SLA studies examining intensive vs. regular instruction (e.g., Author1, 2011; Rogers, 2015), we decided to test students' knowledge right after the treatment had

finished.

We were also interested in examining retention of the words the students learned at the end of the treatment. As the post-test was done at a very short RI (minutes after the treatment), we wanted to make sure the delayed post-test was scheduled at a RI that could be appropriate for the two groups, considering the fact that their treatment was different in terms of ISI (see Rohrer & Pashler, 2007). Following Suzuki and Dekeyser (2017a), this study adopted a 25% ISI to RI ratio, which is considered optimal according to Rohrer and Pashler (2007) and was feasible considering the school calendar. The delayed post-test for the intensive class was thus administered four days after the last session and for the spaced group it was 28 days after the post-test.

Data Analysis

The students' performance with respect to the BVMT in the pre-test, post-test and the delayed post-test was the focus of the study.

In order to investigate language gains between intensive and spaced assisted repeated reading groups, several analyses were performed. Since the BVMT data were normally distributed, parametric tests were used for all data analyses. For the main analysis, a mixed between- and within-subject ANOVA was conducted with time as within-subject factor and group (intensive vs. spaced) as between-subject factor. Post-hoc tests were used to examine the effect of different time intervals (1-day vs. 7-day) on students' vocabulary acquisition and retention. If a significant interaction between group and time was detected in the mixed ANOVA, *t*-tests were later performed for the vocabulary gains from pre-test to post-test, from post-test to delayed

post-test, and from pre-test to delayed post-test for the intensive and spaced groups to examine whether the treatment had differential effects on vocabulary acquisition.

For all analyses, effect sizes were also calculated and were interpreted following Plonsky and Oswald (2014). The authors suggested that for L2 research the benchmarks for between-group contrasts for small, medium and large should be *d* = .40, *d* = .70 and *d* = 1.00 respectively; and for within-group contrasts *d* = .60, *d* = 1.00, and *d* = 1.40. It must be taken into account that these benchmarks are more conservative in interpreting effect size than the traditionally used Cohen's values for small (*d* = .20), medium (*d* = .50) and large (*d* = .80). In the ANOVAs the default index is partial η^2. In this case, we will follow Norouzian and Plonsky (2018), who suggest that the values could be interpreted in the following way: small =.0099; medium = .0588; and large = .1379.

Results

Descriptive statistics for the VLT and BVMT are presented in Table 2. Regarding the VLT scores, the intensive group had an average vocabulary size of 1,921 words while the spaced group had an average size of 1,958 words. The results of the Mann-Whitney *U*-test, used because the scores were not normally distributed, showed there was no significant difference between the intensive and the spaced group (p = .991). To further confirm comparability of groups, we first checked that the initial BVMT scores of the two groups were not statistically different. No significant difference was detected between the two groups (p = .269). The descriptive statistics in Table 2 also show that

the scores of the BVMT post-test were higher than those of the pre-test for both groups, and that in the delayed post-test, the students in the intensive group forgot many of the words they had previously learned.

[Table 2]

The results of the mixed ANOVA (see Figure 3) revealed a significant main effect of time on students' vocabulary learning with a large effect size ($F(2, 68) = 53.114$, $p < .001$, partial $\eta^2 = .61$). The interaction between time and condition was also significant, again with a large effect size ($F(2, 68) = 9.722$, $p < .001$, partial $\eta^2 = .22$), indicating that vocabulary learning was different across conditions. However, there was no significant main effect of treatment on students' vocabulary learning ($F(1, 69) = .119$, $p = .731$, partial $\eta^2 = .002$).

[Figure 3]

To explore further the main effect of time, paired-sample *t*-tests were performed between the three different times, and the *p* value was set to .01 after applying the Bonferroni correction. The results suggest that the scores of the post-test (M = 20.35, SD = 8.89) were significantly higher than those of the pre-test (M = 12.94, SD= 5.19): $t(70) = -10.14$, $p < .001$, indicating significant vocabulary learning after assisted repeated reading, regardless of condition. The estimated effect size was medium ($d = 1.02$) . Moreover, the students' performance in the post-test was significantly better than in the delayed post-test (M = 17.04, SD = 8.63): $t(70) = 4.77$, $p < .001$, which suggests significant vocabulary loss. The effect size of this difference was small ($d =$

-.38). Finally, the difference between pre-test and delayed post-test was also statistically significant: $t(70)= -5.80$, $p < .001$, indicating that, despite the fact that students could not retain many of the words learned during assisted repeated reading, they still experienced significant long-term vocabulary learning. The effect size of this difference was small ($d = .56$).

In order to further examine the interaction between time and treatment, three paired-sample *t*-tests were performed for each group on the vocabulary scores from the pre-test to the post-test, post-test to delayed post-test, and pre-test to delayed post-test separately. After the Bonferroni adjustment, the alpha level was set to .008. Both groups showed significant incidental vocabulary gains after the assisted repeated reading treatment ($t(36) = -8.515$, $p < .001$ and $t(33) = -6.125$, $p < .001$, for the intensive and distributed group respectively). The effect size of the difference was medium for intensive ARR ($d = 1.14$) and small ($d = .87$) for spaced assisted repeated reading. However, from post-test to delayed post-test, the intensive group also showed significant decrease on their vocabulary scores ($t(36) = 5.809$, $p < .001$, $d = -.60$, indicating a medium effect size), which suggests that the students forgot some of the vocabulary they had learned over the treatment period. On the other hand, the spaced group showed no significant loss of learned vocabulary ($t(33) = .722$, $p =.475$, $d = -.06$). Both groups showed significant incidental vocabulary gains from pre-test to delayed post-test ($t(36)= -2.856$, $p =.007$ and $t(33) = -6.044$, $p < .001$, for the intensive and spaced group respectively), indicating that both intensive and spaced assisted repeated reading are effective in promoting incidental vocabulary learning, although the effect size was larger in the case of the spaced group ($d = .82$ vs, $d =.40$).

Additionally, three independent sample *t*-tests were performed to examine how the treatment affected the two target groups differently. The vocabulary gains for intensive and distributed groups were calculated separately from pre-test to post-test, post-test to delayed post-test, and pre-test to delayed post-test (see Table 3 for the descriptive statistics).

[Table 3]

Based on the scores of the pre-test (12.46 and 13.47 out of 36 for the intensive and spaced group respectively), the learners following an intensive schedule had 24 words to learn and those following a spaced schedule had 23 words to learn (out of the 36 included in the test). The scores of the immediate post-test suggest that the learners in the intensive group learned on average 9 words (37.5% of the words that were new for them) while the learners in the spaced group learned 5 words (21.73% of the words they could learn). The results of the inferential statistics indicated there was a significant difference between the gains for the intensive and the distributed group from pre-test to post-test ($t(69)= 2.387$, $p = .020$), indicating that the intensive group incidentally learned more vocabulary than the spaced group in the immediate post-test. The effect size was small ($d = .57$). According to Levene's Test of equality of variances, the assumption of equality of variances was not met ($p = .010$) for the gain score from post-test to delayed post-test. The result showed that the difference in gains between the two groups from post-test to delayed post-test was also significant ($t(60.443) = -4.505$, $p < .001$), indicating that the distributed group retained

significantly more vocabulary than the intensive group during the time from the post-test to the delayed post-test. The effect size of this difference was large (d = -1.05) There was no difference between the gains experienced by the two groups from pre-test to delayed post-test ($t(69)$= -1.476, p =.144, d = -.35). This result confirms that, although assisted repeated reading had more short-term benefits when the practice was concentrated, the longer intersession interval allowed the distributed group to retain their vocabulary gains more successfully. However, in the long run and considering learners' initial knowledge of the target words in the pre-test, both conditions contributed to similar learning outcomes.

Discussion

The current study investigated the effects of assisted repeated reading on vocabulary learning and retention, and how different time distributions of sessions contributed to short- and long-term vocabulary learning. Our results show that both assisted repeated reading with a 1-day interval (intensive) and 7-day interval (spaced) contribute to incidental vocabulary learning. The students knew significantly more words on the post-test than on the pre-test. Moreover, even if significant forgetting took place between post-test and delayed post-test, the scores of the delayed post-test were still significantly higher than in the pre-test, indicating significant long-term vocabulary learning as a consequence of assisted repeated reading. If we look at the effect sizes, the one from pre- to post-test was larger than between the other times, indicating that it is between those two times that more learning occurred. Even when the focus of assisted repeated reading was on comprehension, students in both groups were able to

learn vocabulary. This result is consistent with previous studies showing incidental vocabulary acquisition through assisted repeated reading (Liu & Todd, 2014; Webb & Chang, 2012). However, one limitation of this study was that there was no control group, and there could have been a test-retest effect. Nevertheless, if the results were solely the product of resting, one would expect the results of test 3 (delayed post-test) to be better than those of test 2 (post-test), which was not the case.

We cannot claim, however, that vocabulary gains were exclusively due to assisted repeated reading. It also has to be noted that the comprehension activities used in the study may also have contributed to a certain level of learning but it is hard to avoid the use of such activities if we want to ensure comprehension (Liu & Todd, 2014; Webb & Chang, 2012) or to promote exposure to the target words in a non-explicit form. Another potential factor that could have affected learning is how often students read the words or checked the glossary (which had to be provided because of the high rate of unknown words in the text). Webb and Chang (2015) did not find any significant effect of frequency of exposure on vocabulary learning in assisted reading, so, it might have not been so influential after all, but we cannot discard its potential effect.

As for the second research question, considering the effects of intensive versus spaced assisted repeated reading on incidental vocabulary acquisition, different pictures emerge for learning and retention. Similar to Bird (2010) and Rogers (2015), there was a significant increase from pre-test to post-test for both groups. Yet, when calculating the vocabulary gains from the pre-test to the post-test, our results revealed that the intensive group showed significantly greater vocabulary gains than the spaced group (the intensive group learned almost twice as many words as the spaced: 9 vs. 5).

This result indicates that, when practice is concentrated, it contributes to greater vocabulary learning in the short term (the effect size of the difference between the two groups was small, though). In Bird's or Rogers' study, no significant differences were observed in the post-test for intensive vs. distributed groups in terms of grammar learning. Our findings are thus more in line with previous exploratory research showing an advantage for intensive practice over spaced practice in the short term. Author2 (2007) found that students registered in in both semi-intensive and intensive classes improved significantly more in listening, reading, and writing than the non-intensive class. Author2 (2011) also showed some advantages in lexical richness in written production for intermediate learners attending intensive EFL lessons versus non-intensive ones. It must be taken into account, though, that the apparently higher performance in immediate post-tests in the case of the intensive group could be due to recency effects and to the fact that the time lag between pre- and post-test is shorter in the case of the students following intensive assisted repeated reading than spaced. Nevertheless, this cannot be the whole story, as our results also suggest that the spaced group outperformed the intensive group in the delayed post-test, despite the fact that the lag was longer between the end of the treatment (or the post-test) and the delayed post-test for the spaced group.

Regarding retention, contrary to the advantage for intensive practice for short-term learning, our findings suggest that more distributed practice seems to benefit learners' retention (with a large effect size), when the RI adopted is supposed to be optimal for both groups (Rohrer & Pashler, 2007). This finding is in line with previous research on time distribution (Bird, 2010; Rogers, 2015; Rohrer, 2015; Rohrer

& Pashler, 2007). Similar to Bird (2010), there was a significant decrease from post-test to delayed post-test for the intensive group, with students forgetting about half of the words), but not for the spaced group. In Bird's and Rogers' studies the delayed post-test was administered within what Rohrer and Pashler would consider an optimal ISI/RI ratio for the distributed group but not for the intensive group. What we discovered in our study is that, even when the delayed post-test was administered within the optimal RI also for the intensive group, the lag effect was confirmed.

Although we followed the same ISIs and RIs as in Suzuki and DeKeyser (2017a), we did not find any long-term advantage for the intensive group. Instead, what we found is that the intensive group's scores decreased, approaching those of the spaced group as time progressed (i.e. from post-test to delayed post-test). The divergence between the findings in the current study and Bird's on the one hand, and Suzuki and DeKeyser (2017a) on the other, might be again due to the type of knowledge under investigation, which can be considered declarative in the former two studies and procedural in the latter. Future studies should compare the effect of different time distributions on these two different types of knowledge. Moreover, it would also be interesting to include measures of lexical access speed in future studies examining vocabulary acquisition through assisted repeated reading in intensive vs. spaced schedules in order to check whether similar findings to Suzuki and DeKeyser (2017a) are obtained. The investigation of learners' individual differences for learning under different schedules should also be pursued in future studies, as proposed by Suzuki (2017b), and Suzuki and DeKeyser (2017b).

As our results suggest, with the same amount of time spent on assisted repeated

reading, short-term incidental vocabulary learning is more effective when the time is concentrated but vocabulary learning is better retained when learning sessions are spaced. The spaced group under study retained nearly all the vocabulary gains after 28 days, while the intensive group forgot most of what they had learned after just four days. This finding is consistent with the activation-based model of the spacing effect proposed by Pavlik and Anderson (2005): the activation of the target words in subsequent repetitions of the text in assisted repeated reading for the intensive group was higher than for the spaced group, and this high degree of activation might have prevented their long-term memory encoding. In this sense, the vocabulary gains from the intensive groups might not reflect true learning, since most of the words failed to consolidate into the students' long-term memory.

One limitation of the present study is that, as already reported, the delay between post-test and delayed post-test was not the same for the two groups (4 days vs. 28 days). When examining retention, researchers could try to make the conditions comparable for the intensive and spaced groups by adopting the same delay, or by adopting a delay that has been proposed to be optimal, considering the inter-session spacing. We decided to follow the second path, since many studies have now suggested that depending on when the test takes place concentrated or spaced practice will lead to better results (Bahrick & Hall, 2005; Cepeda et al., 2006). Another possibility was to do two delayed post-tests at two RIs, one supposedly favoring the intensive group and another one the distributed group, according to the suggestions by Rohrer and Pashler (2007). However, we were worried about tiring the students with so many repeated texts and tests.

Implications for teaching and learning

Teachers can use assisted repeated reading since there are more and more learning resources in general and readings in particular that are accompanied by audio. The participants in the present study incidentally learned 37.5% (intensive group) and 21.73% (spaced group) of the words they could learn at an immediate post-test, which is not too different from the 29% in Liu and Todd's study. If our participants had been explicitly asked to learn the vocabulary, they would have probably learned even more words.

If assisted repeated reading is adopted in the classroom, the sessions can be concentrated or spaced depending on the purpose of learning. If the teaching is mainly focused on short-term learning outcomes, it may be better to administer assisted repeated reading sessions in a more concentrated distribution. For better retention of learning, on the other hand, teachers can consider distributing the sessions using a 7-day intersession interval. Since reading resources with audio are abundant, learners who want to study on their own can use assisted repeated reading sessions in a similar way—concentrated assisted repeated reading for immediate vocabulary learning or repeatedly reading the same passage with longer intervals between each reading to retain learned words longer.

Acknowledgments

This research was supported by the Spanish Science Ministry Grant FFI2016-80576-P as well as the Catalan Research Agency Grants 2014SGR1089 and 2017SGR560.We

would like to thank the anonymous reviewers as well as the editors Ahmar Mahboob and Charlene Polio for their guidance and insightful comments.

Authors Bio

Raquel Serrano is an Associate Professor in the Department of Modern Languages and Literature and English Studies at the University of Barcelona. Her research focuses on the role of time distribution and learning context in the acquisition of English as a second language.

Hsiao-yun Huang is an English teacher in Da-tung senior High School in Taiwan. She has an M.A. in Applied Linguistics and Language Acquisition in Multilingual Settings from the University of Barcelona. She does research on vocabulary acquisition, focusing on Taiwanese learners of English.

References

Bahrick, H.P., & Hall, L.K. (2005) The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52,* 566–577. http://dx.doi.org/10.1016/j.jml.2005.01.012

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31,* 635–650. https://doi.org/10.1017/S0142716410000172

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research

and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354, 380. http://dx.doi.org/10.1037/0033-2909.132.3.354

Chang, A., & Millett, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language*, *25*(2), 126–148. https://doi.org/10.1016/j.system.2015.05.003

Chou, C., Chiu, M., Hong, M., Chen, C., Chen, P., Huang, Y., …Hall, S. (2010). *High school English reader*. Taipei, Taiwan: Lungteng.

Collins, L. and White, J. (2011) An intensive look at intensity and language learning. *TESOL Quarterly, 45,* 106-133. http://dx.doi.org/10.2307/3587881

Collins, L., Halter, R. H., & Lightbown, P. M. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly, 33*(4), 655-680. https://doi.org/10.2307/3587881

DeKeyser, R. (1995). Learning Second Language Grammar Rules: An Experiment With a Miniature Linguistic System. *Studies in Second Language Acquisition, 17*(3), 379-410. https://doi.org/10.1017/S027226310001425X

Gorsuch, G. (2015). Repeated reading for Japanese language learners : Effects on reading speed, comprehension, and comprehension strategies. *The Reading Matrix: An International Online Journal*, *15*(2), 18–44. Retrieved from http://www.readingmatrix.com/

Han, Z., & Chen, C. A. (2010). Repeated-reading-based instructional strategy and

vocabulary acquisition : A case study of a heritage speaker of Chinese. *Reading in a Foreign Language, 22*(2), 242–262. Retrieved from http://nflrc.hawaii.edu/rfl/

Hulstijn, J.H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 2632-2637). Malden, MA: Blackwell Publishing Ltd. http://dx.doi.org/10.1002/9781405198431.wbeal0530

Laufer, B. (2006). Comparing focus on form and focus on formS in second-language vocabulary learning. *Canadian Modern Language Review, 63,* 149-166. http://dx.doi.org/10.3138/cmlr.63.1.149

Liu, Y. T., & Todd, A. G. (2014). Implementation of assisted repeated reading techniques for the incidental acquisition of novel foreign vocabulary. *Language Teaching Research*, *20*(1), 53-74. https://doi.org/10.1177/1362168814559802

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge : Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82. http://dx.doi.org/10.3138/cmlr.63.1.59

Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research, 34,* 257-271. https://doi.org/10.1177/0267658316684904

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29,* 559-586. http://dx.doi.org/10.1207/s15516709cog0000_14

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38,* 97-130. http://dx.doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22,* 31-55. Retrieved from http://nflrc.hawaii.edu/rfl/April2010/articles/pellicersanchez.pdf

Peters E., Hulstijn J., Sercu L., Lutjeharms M. (2009). Learning L2 German vocabulary through reading : the effect of three enhancement techniques compared. *Language Learning, 59* (1), 113-151. http://dx.doi.org/10.1111/j.1467-9922.2009.00502.x

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64,* 878-912. http://dx.doi.org/10.1111/lang.12079

Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Philadelphia, PA: John Benjamins. http://dx.doi.org/10.1075/lllt.2.08rob

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition,27* (2), 235–268. http://dx.doi.org/10.1017/S0272263105050126

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly, 49, 857-866.* http://dx.doi.org/10.1002/tesq.252

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition.

Applied Linguistics 2017. http://dx.doi.org/10.1093/applin/amw052

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, *27*(4), 635–643. https://doi.org/10.1007/s10648-015-9332-4

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Psychological Science*, *16*(4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Rohrer, D., & Taylor, K. (2006) The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224. http://dx.doi.org/10.1002/acp.1266

Samuels, S. (1979). The method of repeated readings. *The Reading Teacher, 32*(4), 403-408. Retrieved from http://www.jstor.org/stable/20194790

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke : Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviours of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–88. https://doi.org/10.1177/026553220101800103

Serrano, R. (2011). The time factor in EFL practice. *Language Learning, 61*, 117-145.

Serrano, R. (2012). Is time concentration good or bad for learning? Reflecting on the results from the cognitive psychology and from the SLA literature. In C. Muñoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 3-22). Clevedon: Multilingual Matters.

Serrano, R., Muñoz, C. (2007). Same hours, different time distribution: Any difference

in EFL? *System 35*, 305-321. http://dx.doi.org/10.1016/j.system.2007.02.001

Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programs. *Language Teaching Research*, *19*, 89-106. https://doi.org/10.1177/1362168814541748

Suzuki, Y. (2017a). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning*, *67*, 512-545. http://dx.doi.org/10.1111/lang.12236

Suzuki, Y. (2017b). The role of procedural learning ability in automatization of L2 morphology under different learning schedules: An exploratory study. *Studies in Second Language Acquisition.* http://dx.doi.org/10.1017/S0272263117000249

Suzuki, Y., & DeKeyser, R. (2017a). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research,21(2),*166-188. https://doi.org/10.1177/1362168815617334

Suzuki, Y., & DeKeyser, R. (2017b). Exploratory research on L2 practice distribution: An aptitude-by-treatment interaction. *Applied Psycholinguistics, 38,* 27-56. https://doi.org/10.1017/S0142716416000084

Webb, S., & Chang, A. (2012). Vocabulary Learning through Assisted and Unassisted Repeated Reading. *Canadian Modern Language Review*, *68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686. https://doi.org/10.1177/1362168814559800

Table 1

A summary of studies investigating different time distribution in L2 learning

Study	Learning items	Inter-session Interval (ISI)	Retention Interval(RI)	Results
Rohrer & Pashler (2007)	(1) names of obscure objects (2) Swahili-English word pairs	Varying ISIs: (1) 5-min~14 day (2)5-min~6-month	(1) 10-day (2) 6-month	(1) 1-day ISI yielded best recall (2) optimal ISI roughly 1 month
Bird (2010)	English syntax	5-day 14-day	7-day 60-day	No difference at 7-day RI; distributed group outperformed concentrated group at 60-day RI
Rogers (2015)	English L2 grammar learning	2.25-day 7-day	42-day	Significant advantage for the spaced group at long RI
Suzuki & DeKeyser (2017)	Japanese morphological structure	1-day 7-day	7-day 28-day	Concentrated group outperformed distributed group at descriptive level; in long RI, significant interaction in favor of the concentrated group
Suzuki (2017a)	Morphology in Supurango (novel miniature language)	1-day 3.3-day	7-day 28-day	Concentrated group outperformed distributed group in accuracy at both RI. No differences in speed

Table 2

Descriptive statistics: VLT and BVMT

	Intensive (n = 37)	Spaced (n = 34)
VLT /3000	1921.14 (356.96)	1957.68 (265.96)
BVMT Pre-test /36	12.46 (5.35)	13.47 (5.06)
BVMT Post-test /36	21.49 (9.85)	19.12 (7.68)
BVMT Delayed post-test /36	15.57 (9.57)	18.65 (7.29)

Table 3

Results of students' word gains

	Intensive (n = 37)	Spaced (n = 34)
Gains Pre-Post	9.02 (6.45)	5.65 (5.38)
Gains Post-Delayed	-5.92 (6.19)	-0.47 (3.79)
Gains Pre-Delayed	3.10 (6.61)	5.17 (4.99)

	Session 1		Session 2		Session 3		Test
1-day ISI	Monday	1 day	Tuesday	1 day	Wednesday	10 days	ISI=10% RI
7-day ISI	Monday	7 days	Monday	7 days	Monday	10 days	ISI = 70% RI

Figure 1. *Examples of different intersession intervals (ISI) and retention intervals (RI)*

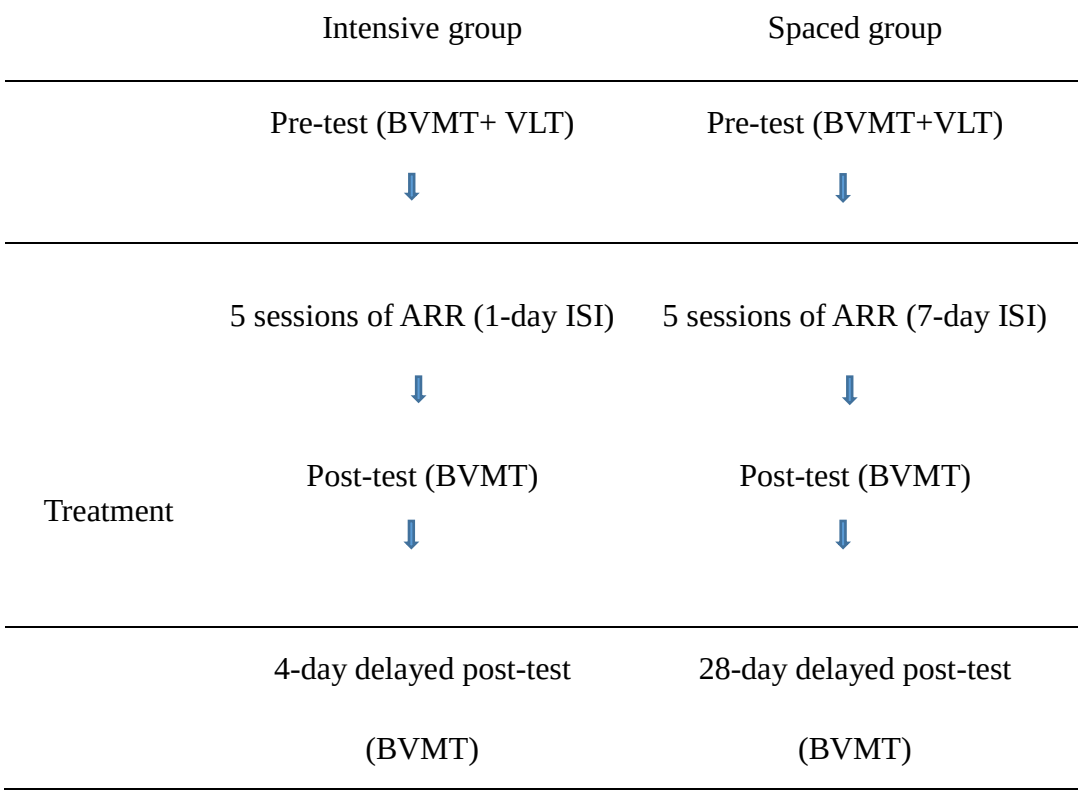

Figure 2. *Research design*

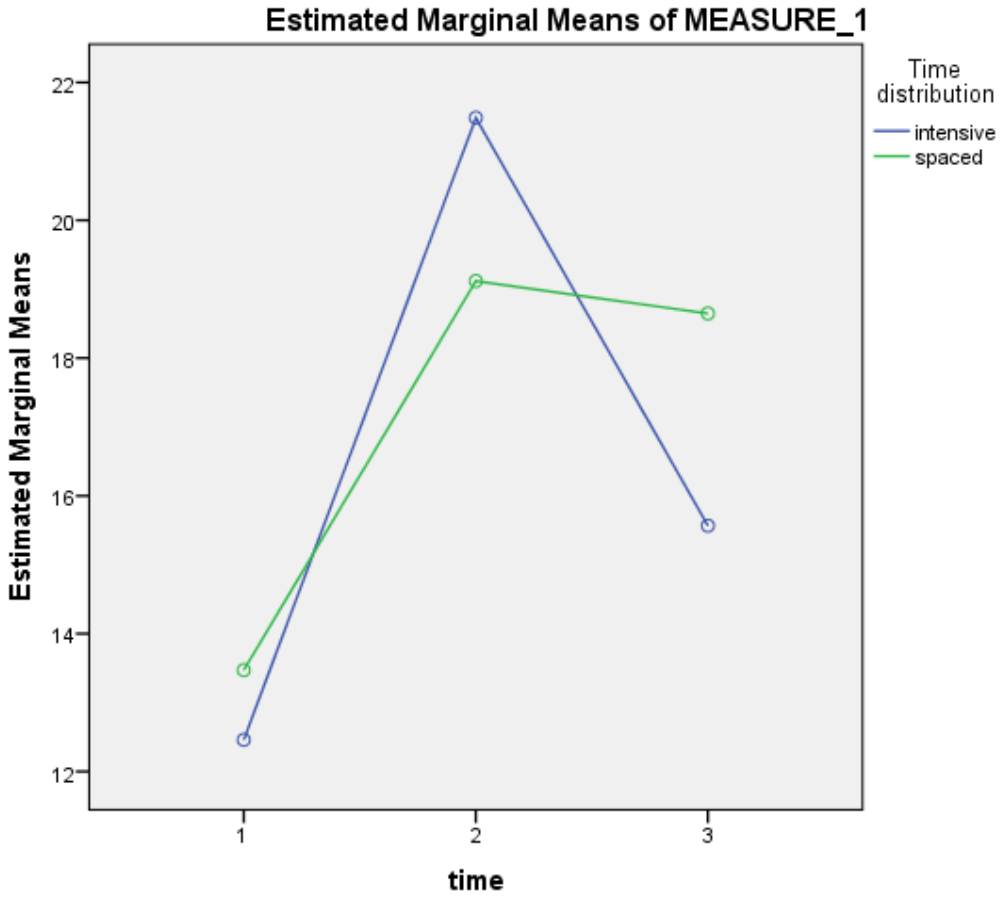

Figure 3. *BVMT scores pre-test (1), post-test (2) and delayed post-test (3)*