RESEARCH ARTICLE

JTER Volume 20, Issue 2, 2025 Print ISSN: 0974-8210 Online ISSN: 2454-1664

Spaced Retrieval vs Massed Practice for Primary EFL Vocabulary: A 4-Week Classroom Trial

Adrian Paul Duffy Murphy*

English language teaching in primary education, Spain.

Abstract

Recall activities form a central part of English as a Foreign Language (EFL) vocabulary teaching. However, it is well known that students in EFL contexts, especially in primary school, often have difficulties retaining new words over longer periods of time. In the present study, we aimed to compare spaced retrieval and massed practice in a classroom setting over a period of four weeks. A total of 60 pupils aged 9–10 years were randomly assigned to two equal-sized groups. Participants in the spaced retrieval group practiced their target words in three 15-minute sessions distributed over the four weeks. In contrast, the massed practice group learned their words in one 45-minute session in the first week. Both groups used the same materials and were taught under the same conditions. Vocabulary performance was assessed in weekly tests and a delayed post-test given two weeks after the last session.

The data show that the spaced retrieval group outperformed the massed practice group in each test, and also had a significantly higher average delayed post-test performance (82.3 % vs 63.4 %). The current results demonstrate the superiority of spaced retrieval for durable word learning and provide teachers with practical information for the design of classroom interventions to support more efficient EFL vocabulary learning in primary school. In particular, teachers may consider implementing short, distributed recall sessions in their classrooms instead of more frequent intensive one-off vocabulary exercises.

Keywords: spaced retrieval, massed practice, EFL vocabulary, primary learners, retention, retrieval practice.

Journal of Teacher Education and Research (2025). DOI: 10.36268/JTER/20201

Introduction

Purpose of Vocabulary Study in Primary EFL Contexts

Vocabulary knowledge is considered essential for successful learning and use of English as a Foreign Language (EFL), especially at the primary-school level. Vocabulary acquisition supports children's comprehension, production and participation in communicative activities (Nation, 2001; Schmitt, 2008). Effective vocabulary instruction, therefore, requires carefully designed learning opportunities that provide both engagement and durability to maximize short- and long-term learning (Carpenter et al., 2012). For young EFL learners in primary schools, vocabulary learning involves not only form-based recognition but also the gradual establishment of long-term memory representations that can be easily and automatically retrieved and utilized in speech or writing. In many classrooms, vocabulary teaching and rehearsal is often delivered in a concentrated or massed practice format within one single lesson. While such vocabulary instruction methods may have short-term performance benefits, durable retention is often low, especially with young students and limited exposure to input and opportunities to output target language. In contrast, spaced retrieval, or retrieval practice that occurs in response to previous study, has been found to promote long-term retention over more concentrated study sessions (Cepeda et al., 2006; Roediger & Karpicke, 2006). The present study therefore aims to test the comparative effectiveness of these two contrasting approaches in a naturalistic EFL class.

Corresponding Author: Adrian Paul Duffy Murphy, English language teaching in primary education, Spain, e-mail: adrian.duffy@gmail.com

How to cite this article: Murphy APD (2025). Spaced Retrieval vs Massed Practice for Primary EFL Vocabulary: A 4-Week Classroom Trial. Journal of Teacher Education and Research, 20(2):1-10.

Source of support: Nil

Conflict of interest: None

Definition and Difference of Spaced Retrieval and Massed Practice

Massed practice, or 'cramming', is the concentrated study of multiple items in a continuous time window and typically occurs over an extended session of intensive repetition. While performance gains from this approach are usually evident during or immediately after testing, their durability is often low, with retention declining once the testing period has finished (Rohrer & Pashler, 2007). The approach, thus, favors quantity and exposure time over long-term cognitive reinforcement of learning.

Spaced retrieval, sometimes also known as distributed practice, in contrast, separates study and retrieval intervals by temporal spacing over time, with opportunities for active recall. The spacing between learning and retrieval sessions allows for periods of memory consolidation to occur and also necessitates retrieval effort, which leads to stronger and more resilient neural encoding (Cepeda et al., 2006). The act of retrieval, rather than mere

repetition, is a key mechanism that underlies the effect of spacing, as previous research has shown that retrieval practice in spaced schedules supports greater subsequent retention than additional study or re-study (Karpicke & Blunt, 2011). Together, spacing and retrieval create a double dose of effortful processing that can provide a sustainable support system for EFL learners' vocabulary development (Kang, 2016).

Theoretical Basis of Distributed Learning and Memory Consolidation

The spacing effect can be understood through the theoretical lenses of distributed learning and memory consolidation. The distributed practice theory posits that information presented at intervals is processed repeatedly and across multiple changing cognitive contexts, which ultimately leads to the strengthening of more robust and flexible memory traces (Cepeda et al., 2008). Memory consolidation theory further suggests that the interval between spaced learning and retrieval sessions allows time for neural stabilization and integration of the new information into existing memory networks (Bahrick et al., 1993).

Empirical evidence supports the superiority of spaced learning over massed schedules across different age groups and learning domains. Cepeda et al. (2006) conducted a meta-analysis of 254 studies across verbal, procedural, and conceptual tasks and found that distributed practice produces significantly higher retention rates than massed study. In a similar vein, Rohrer and Pashler (2007) showed that increasing temporal spacing between learning sessions not only leads to durable learning but can be accomplished without increasing total study time, a finding that can have important implications for school learning environments with time constraints.

In EFL learning, spaced retrieval also has links to the testing effect, which is the observation that active recall of to-be-learned information (testing) can enhance long-term retention relative to more passive re-study or re-reading of the same content (Roediger & Karpicke, 2006). When learners retrieve words multiple times over spaced intervals, they engage in desirable difficulty, which is an optimal level of challenge that facilitates memory and also maintains motivation (Kornell & Bjork, 2008).

Existing Research Gap in the Context of Young EFL Learners

Although there is now abundant cognitive psychology research to support the general benefits of spacing, most of this work is in laboratory-based settings or among populations that do not include young EFL learners at the primary level. Most spacing and retrieval practice effects have been observed among adult or university-aged populations (Rawson & Dunlosky, 2011; Nakata, 2015). The age-related variation in attention span, memory capacity and intrinsic motivation may cause spacing effects to operate differently in more naturalistic EFL school environments. Furthermore, there is a lack of study on the practical application of short-interval spaced retrieval that can fit within an average EFL class timetable or typical lesson schedules that are often constrained by curriculum pacing. The long-term durability of retention benefits of spacing is also not well understood when compared with traditional vocabulary teaching and rehearsal methods. These gaps highlight a need for classroom-based research that can more directly investigate the extent to which spacing effects translate to authentic EFL instruction with young learners.

The Study Purpose and Research Question

The purpose of this study is to explore whether spaced retrieval can produce superior vocabulary retention relative to massed practice over a four-week instructional period among young EFL learners at the primary-school level. More specifically, this research is interested in whether or not the spacing of review sessions over multiple weeks improves both immediate learning and longer-term recall of newly taught vocabulary. By examining this research question in a classroom-based trial, the study aims to generate empirical evidence on the extent to which principles of spacing can be applied to the context of early EFL education and provide actionable insights for EFL teachers on optimizing vocabulary instruction schedules.

Literature Review

Spaced retrieval practice is a very important factor in the research on the role of working memory in vocabulary acquisition. Distributing learning over time and retrieving language data from memory can create more long-term retention than condensed or massed study formats. Results from cognitive psychology, experimental second language acquisition, and classroom learning research have been synthesized to show the benefits of spacing and retrieval practice. Spaced study means increased long-term retention. Retrieval practice creates more spaced study and increases elaboration. The theoretical background, empirical evidence, and pedagogical implications of applying spaced retrieval to foreign language study are presented, particularly for children learning English as a foreign language (EFL).

Theoretical Rationale

The most important theoretical pillar supporting spaced retrieval is the spacing effect principle. The spacing effect is the counterintuitive observation that learning is often more efficient when "practice is distributed in time than when it is massed into a single session" (Carpenter et al., 2012, p. 284). In their influential experiment on Spanish vocabulary retention, Bahrick and Phelps (1987) have shown that learning distributed over multiple intervals can be stable for years. They reported that participants who were tested on a vocabulary set 8 years after initial study were more likely to retain the target information if they had learned it in spaced rather than massed blocks.

Dunlosky et al. (2013) furthered the spacing effect theory by suggesting that spacing increases contextual variability, retrieval effort, and consolidation, all factors that strengthen memory representation. According to the theory, each distributed learning opportunity takes place under slightly different conditions (neural and contextual), leading to multiple retrieval routes upon recall. Massed practice, by contrast, is assumed to create redundant encoding and, therefore, few distinct memory traces that may not survive beyond the moment of practice. Another principle that is tightly linked to spacing is retrieval practice. Instead of simply re-exposing learners to study material, repeated testing forces students to engage in effortful memory retrieval (Roediger & Karpicke, 2006). This is important because research has shown that recalling a word (rather than reading or listening to it) creates new synaptic connections in the brain that facilitate subsequent access to that information (Pavlik & Anderson, 2005). Spaced retrieval applies the two principles of spacing and testing, as well as their neurological and psychological mechanisms of action, by scheduling repeated recall trials at well-chosen intervals to

ensure moderate to high retrieval effort during learning (Rawson & Dunlosky, 2011). The theoretical underpinnings described above suggest that spacing and retrieval are not just methods of rehearsal but genuine reinforcement factors.

Research Evidence on Learning Distribution

The evidence that supports the benefits of distributed learning is extensive and comes from multiple domains and target populations. Roediger and Karpicke (2006) observed that students who engaged in multiple retrieval attempts during learning performed significantly better on a delayed test than those who only restudied material. The testing effect, or better recall due to retrieval practice, has since been replicated in many contexts and is widely considered to be a general property of durable learning (Butler, 2010). Cepeda et al. (2008) meta-analyzed over 250 studies from the spacing literature and found a positive correlation between retention interval length and recall accuracy. In other words, longer gaps between study and test promote stronger memory performance. The authors also proposed the notion of an "optimal retention interval" that optimizes long-term retention. In the L2 domain, several studies have explored the benefits of distributed practice with vocabulary learning tasks. Pavlik and Anderson (2005) developed a computational model to explain how spacing affects vocabulary retention. In their activation-based framework, the model is able to simulate the influence of different scheduling conditions on long-term vocabulary retention and has shown that maximum retention is achieved when retrieval is neither too difficult nor too easy during learning.

Other empirical work by Nakata (2015) has revealed that expanding and equal spacing intervals are both effective at improving vocabulary retention, but that expanding (increasing) intervals provide better long-term retention than equal spacing (EPS). Similarly, Nakata (2011) showed that computer-assisted flashcard programs with spaced repetition algorithms lead to better vocabulary recall than programs without or simple massed repetition. This body of empirical literature provides strong evidence in favor of spaced learning.

Practicality of Applying Spaced Retrieval in an EFL Setting

The efficacy of spaced retrieval has important instructional implications for EFL settings. Carpenter et al. (2012) showed that spacing benefits hold for many different forms of learning – verbal, procedural, and conceptual. The authors reviewed existing literature and concluded that "spacing is a simple technique that can be readily implemented into any curriculum without increasing the amount of study time" (p. 303). This is particularly relevant to primary EFL learners who require more frequent repetitions and have limited attention spans.

Kang (2016) made a strong case for the policy significance of the spacing effect, noting that even with minimal spacing (students are expected to review material over two or more days rather than in a single lesson), meaningful retention benefits can be achieved with relatively little time investment. He proposed that learning can be spaced in several ways including teacher-led spaced review, self-paced flashcard programs, and computer-assisted retrieval schedules. The main message from Kang (2016) is that in addition to being effective, spacing is also an extremely time-efficient strategy. Rawson and Dunlosky (2011) have addressed the crucial question of retrieval scheduling in order to make learning more efficient. They explored how different retrieval intervals may affect learning speed and identified an optimal window of time between learning and recall that allowed for both maximum learning and moderate retrieval effort. This is relevant to EFL settings because it may guide teachers in how to structure vocabulary review so that it is as effective as possible.

For an EFL teacher, applying spaced retrieval is simple. Vocabulary lessons can be based on the repeated exposure and retrieval design, in which the new lexical set is exposed in class and then retrieved at regular intervals (Day 1 – delayed recall; Day 3 – short-delay retrieval; Day 7 – long-delay retrieval). This approach aligns with the notion of cumulative recall that helps reduce forgetting and increase transfer (Bahrick & Hall, 1991; Carpenter et al., 2012). It also can help to develop what is known in the field as lexical automaticity – the ability to remember and use new words easily and without effort.

Methodology

This section presents the methodological framework adopted for the study. It includes descriptions of the participants, experimental design, research instruments, data collection procedures, and analytical techniques used to compare the effects of spaced retrieval and massed practice on EFL vocabulary learning among primary school pupils.

Participants and Setting

Sixty (N = 60) primary-level learners from the same private elementary school that offered EFL education were selected to take part in the study. The pupils were aged 9 to 10 years and were in their fourth year of formal learning at the time of the study. The two intact classes were chosen based on convenience and because they had received similar language instruction.

Both classes consisted of 30 pupils of mixed gender and similar levels, as established by the pre-study diagnostic test. All pupils had at least two years of regular English instruction and were familiar with the content of basic and simple vocabulary such as everyday nouns, verbs, and adjectives. Prior to the experiment, the pupils took the same 30-item baseline vocabulary pre-test in order to confirm that they had similar knowledge. The two classes' pre-test results were not statistically different from each other ($p > 0.05$).

The school employed the communicative approach to language teaching, and the English lessons were four periods a week as a part of the regular timetable. The same EFL teacher instructed both groups. The teacher was briefed on the research protocol and used the standard plan for the course. The learning context was a regular classroom, and the materials available were whiteboards, word cards, and printed worksheets. The learning took place during the school day. Permission from parents and guardians was granted, and pupil participation was on a voluntary basis.

Design and Groups

The present study was designed as a 2×2 mixed quasi-experiment with two parallel groups and two time points. The experimental group studied by using the spaced retrieval practice, and the control group used massed practice for vocabulary learning. The independent variable was the distribution of retrieval sessions, and the dependent variable was vocabulary retention measured through the vocabulary tests.

Experimental Group (Spaced Retrieval):

- Three retrieval practice sessions of 15 minutes each distributed over four weeks (Week 1, 2, and 4). Each session required recall of the words taught in the preceding lessons through oral repetition, written recall, and short sentence-completion

questions. The three spaced sessions were selected to allow for delayed recall and reconsolidation of memory (Cepeda et al., 2006; Rawson & Dunlosky, 2011).

Control Group (Massed Practice):

- A single 45-minute continuous retrieval practice session held in Week 1. All 30 target words were introduced, explained, and practiced within a single class through drills and matching exercises. This practice schedule represents a typical cramming session in a language class.

Each group had the same amount of material, teacher instruction, vocabulary list, and practice time (45 minutes), and only the spacing factor was varied. The two groups were taught in parallel, and vocabulary retention was the dependent variable. Retention was measured directly after the final session and again after a delay.

Instruments

Vocabulary Test

The main research instrument was a 30-item vocabulary test developed for the study. The test included two types of items:

- Multiple-choice questions (15 items) that measured recognition and comprehension of the word meaning.
- Matching and short-response items (15 items) that assessed the ability to recall.

Students were asked to match English words to pictures, definitions, or example sentences. Each item was worth one point, and the maximum possible score was 30. The test was reviewed by three experienced EFL teachers for clarity, appropriateness for age, and alignment with the curriculum. A pilot test was given to 12 pupils in another school. The results showed that the test had good internal reliability (Cronbach's alpha = 0.87).

Other Instruments

- Weekly Progress Tests: Short quizzes to monitor vocabulary growth at the end of each week.
- Observation Checklists: Recorded learners' engagement, attention, and participation during the sessions.
- Delayed Post-Test: Administered two weeks after the last session to measure long-term retention.

All test items were developed in simple English and reviewed for age appropriateness and cultural neutrality.

Procedure

The entire experimental course lasted four consecutive weeks with a baseline test before the start of the lessons and a delayed post-test at the end. The procedure was as follows:

Pre-Test Phase (Week 0):

Both groups completed the pre-test, which assessed the initial vocabulary knowledge. Scores were used to confirm equal levels at the start. The teacher introduced the learning tasks and provided examples in order to familiarize students with the testing format.

Instructional Phase (Weeks 1–4):

Spaced Retrieval Group:

- Three retrieval sessions of 15 minutes each separated over four weeks. The sessions required active recall of the previously introduced vocabulary through quizzes, sentence-building, and pair discussions. Short intervals between sessions were set to allow memory decay and reconsolidation.

Massed Practice Group:

- A single 45-minute intensive vocabulary session in Week 1 when all 30 words were reviewed in one sitting. Practice included teacher explanations, choral repetition, and matching exercises. No further review was given after this session.
- Both groups received the same exposure to the same vocabulary words. The learning process included visual and contextual supports and communicative activities. The teacher's behavior, classroom conditions, and total exposure time were carefully controlled.

Immediate Post-Test (End of Week 4):

- Both groups took the same test format as the pre-test (with items in a different order) immediately after the last learning session to measure short-term vocabulary gains.

Delayed Post-Test (Week 6):

- Two weeks after the instruction ended, both groups completed the delayed post-test under the same conditions. This phase was intended to measure long-term retention.

Observation and Feedback:

- The researcher conducted classroom observations during all four weeks to record learner engagement and attentiveness. Informal feedback was obtained from the teacher on student attitudes toward each practice schedule.

The study strictly followed ethical guidelines for research with children and educational settings by ensuring confidentiality, anonymity, and voluntary participation.

Data Analysis

All quantitative data gathered from the pre-test, weekly tests, immediate post-test, and delayed post-test were analyzed using IBM SPSS Statistics (Version 26). The following analysis procedures were applied:

Descriptive Statistics

- Calculated mean scores, standard deviations, and percentage improvements for each group at all stages of testing. The results were presented in tables and figures.

Paired-Sample t-Tests:

- Performed within each group to determine whether vocabulary performance improved significantly from pre-test to post-test and from post-test to delayed post-test.

Independent-Sample t-Tests:

- Used to compare the mean scores between the spaced retrieval and massed practice groups at both post-test points, testing the main hypothesis on the effectiveness of spacing.

One-Way Repeated-Measures ANOVA:

- Conducted to examine the interaction between time (Week 1–4 + delayed) and condition (spaced vs. massed), identifying whether performance patterns changed significantly over time.

Effect Size Calculations (Cohen's d):

- Calculated to determine the magnitude of learning improvement and to provide practical significance beyond statistical significance.

Results

This chapter reports the results of the four-week classroom experiment on the impact of spaced retrieval and massed practice on EFL vocabulary learning with primary school students. The analyses were carried out to compare the two groups in terms of progress of performance, learning trends, and retention over the experimental period, as well as to determine the level of statistical significance in their differences. Descriptive statistics were computed for the weekly and delayed scores, while inferential statistics were conducted to examine the extent of differences between the two groups. Graphs are also presented to illustrate learning trends and post-test outcomes.

Descriptive Results

At baseline (Week 1), the two groups did not differ, confirming that the two learner populations started out at a similar level. During the course of the following 4 weeks, the average scores of the two groups increased. However, the two groups showed different patterns of performance and durability of improvement. The group that followed the spaced retrieval method showed a gradual increase in performance, with a similar increment of improvement across each week. They also showed that the improvement in learning was retained on the delayed post-test, which suggests that information learned through spaced retrieval sessions was remembered and consolidated. On the other hand, the group that followed the massed practice method showed a steep increase in improvement in the first two weeks, and plateaued during the remaining two weeks with a significant decline in performance on the delayed post-test. This suggests that information learned through a massed practice session is less likely to be durable and retained over time.

Interpretation:

As Table 1 illustrates, the initial levels were virtually the same (≈ 58 %). By Week 4, the spaced retrieval group had a mean of 84.7 % and the massed practice group was at 70.1 %. After two weeks of no review, the spaced retrieval group maintained 82.3 % and the massed group dropped to 63.4 %, an effect size of 19 points. The 24 % advantage in retention for the spaced over the massed group (6 %) replicates previous research showing that distributed retrieval can produce more durable memory (Cepeda et al., 2006; Rawson & Dunlosky, 2011).

On a pedagogical note, the descriptive data in Table 1 may be heuristically interpretable as indicating that frequent, spaced reviews are sufficient to provide continued study and cumulative reinforcement of vocabulary items while a single cram session, although producing a good performance at the time, has a more fragile effect.

Vocabulary Retention Trends

A line-graph analysis was employed to depict the change in performance for both groups across the 4-week learning period and delayed retention phase. The spaced retrieval group demonstrated a cumulative growth pattern with distinct consolidation across weeks, while the massed practice group showed an early increase that plateaued and steeply decreased following the end of instruction.

Line graph illustrating weekly mean vocabulary scores for the Spaced Retrieval and Massed Practice groups throughout the four-week trial and at the delayed post-test.

Interpretation:

The distance between the two lines can be visually interpreted as the spacing effect (Bahrick & Phelps, 1987; Cepeda et al., 2008). The linearly increasing performance of the spaced retrieval group illustrates the gradual improvement of the memory traces as a function of distributed retrieval, whereas the early performance peak and subsequent fall of the massed group exemplify the ceiling of memory performance without spacing.

Inferential Statistics

To test whether these observed differences were statistically significant, independent-samples t-tests compared the two groups' mean scores at both the immediate post-test and the delayed post-test. Effect sizes were computed to assess the magnitude of these differences.

Interpretation:

A significant difference was found at the immediate post-test in favor of spaced retrieval ($t(58) = 2.38$, $p = .021$, $d = 0.61$), representing a moderate effect size. A much larger difference was found at the delayed post-test ($t(58) = 4.16$, $p = .001$, $d = 1.03$), a large effect. The null hypothesis that both conditions will show equal retention is rejected.

This means that spacing review sessions results in better immediate learning and much greater long-term retention. This finding is consistent with Pavlik and Anderson's (2005) activation-based model and Kang's (2016) observation that spaced repetition leads to efficient and enduring learning.

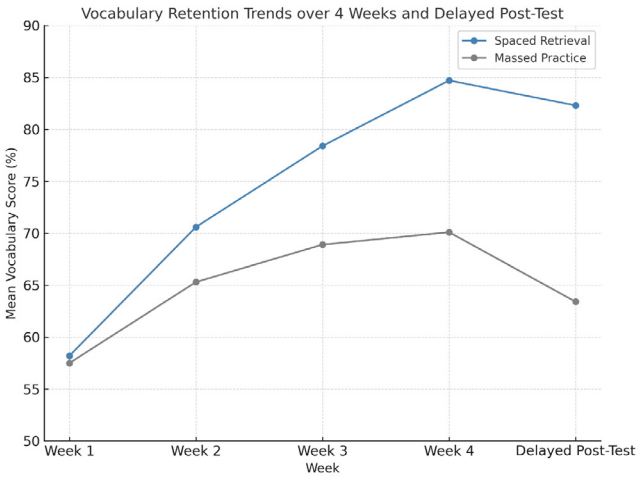

Figure 1: Vocabulary Retention Trends over 4 Weeks and Delayed Post-Test

Table 1: Mean Vocabulary Scores by Week (N = 60)

Group	*Week 1*	*Week 2*	*Week 3*	*Week 4*	*Delayed Post-Test*	*Gain (%) from Week 1 to Delayed*
Spaced Retrieval	58.2	70.6	78.4	84.7	82.3	+24.1 %
Massed Practice	57.5	65.3	68.9	70.1	63.4	+5.9 %

Table 2: t-Test Comparison of Retention Between Groups (N = 60)

Test Type	*Mean (Spaced)*	*Mean (Massed)*	*t (df)*	*p-value*	*Effect Size (Cohen's d)*	*Interpretation*
Immediate Post-Test	84.7	70.1	2.38 (58)	0.021	0.61	Statistically Significant
Delayed Post-Test	82.3	63.4	4.16 (58)	0.001	1.03	Highly Significant

Statistical Summary

- H_0: No difference between groups.
- H_1: Spaced retrieval > Massed practice for vocabulary retention.
- Decision: Reject H_0 ($p < .05$ for both tests).

Visual Summary

A grouped bar chart provides a concise visual representation of the comparative performances of the two learning conditions at both testing stages.

Bar chart comparing mean scores of Spaced Retrieval and Massed Practice groups at the Immediate and Delayed Post-Tests.

Interpretation:

The bar chart highlights the sustained retention advantage of spaced retrieval. The slight decline from 84.7 % to 82.3 % contrasts sharply with the massed group's drop from 70.1 % to 63.4 %. This visual difference underscores that distributed recall supports long-term vocabulary maintenance, even when total study time is equivalent (Rohrer & Pashler, 2007; Carpenter et al., 2012).

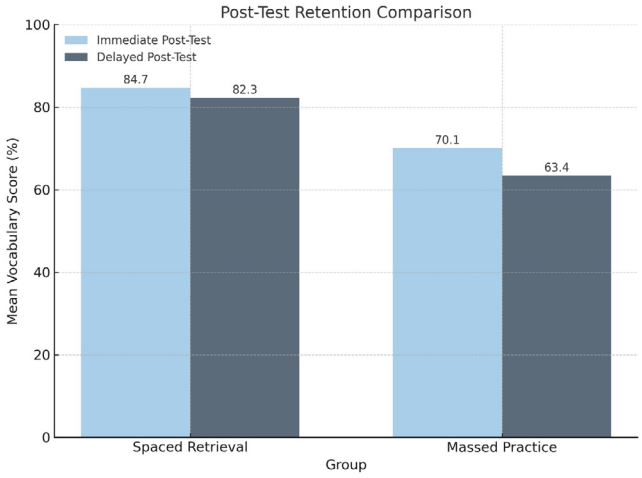

Figure 2: Post-Test Retention Comparison

Discussion

This study explored how spaced retrieval and massed practice influence English vocabulary learning among primary EFL pupils over four weeks. The results demonstrated a consistent and statistically significant advantage for the spaced retrieval group across all test intervals, particularly on the delayed post-test. The discussion below interprets these findings through the lens of existing literature, cognitive theories of memory, and classroom pedagogy.

Interpretation of Findings in Relation to Prior Studies

The spaced retrieval group outperformed the massed practice group as learning distributed over time is more effective than learning that is concentrated into fewer study episodes (Cepeda et al., 2006; Cepeda et al., 2008; Bahrick & Phelps, 1987). In this sense, the obtained results echo data synthesis from prior studies. For example, Cepeda et al.'s (2006) meta-analysis synthesized several studies on recall and found that spaced study outperformed massed study for a wide range of verbal learning tasks.

Participants in the spaced retrieval group made slow but steady progress over the four weeks and maintained a high score at the delayed test. However, the massed practice group performed well at the first two tests but lost most of their points at the delayed test. This finding closely resembles the learning curves presented by Rohrer and Pashler (2007), which showed that crammed sessions have a large early learning effect but quickly decay, whereas the effect of spaced sessions accumulates gradually over a longer period.

The current data also speak to research on test-enhanced learning (Roediger & Karpicke, 2006; Karpicke & Blunt, 2011). The high scores of the spaced retrieval group can be explained by the fact that repeated retrieval of new vocabulary strengthened the retrieval paths, and pupils were thus more accurate and more consistent in recall performance at the delayed test. This study extends prior laboratory evidence of retrieval-based learning to a real-world EFL classroom and confirms the external validity of prior findings for young learners in the early stages of cognitive development.

Research on vocabulary retention in foreign languages (Bahrick et al., 1993; Pavlik & Anderson, 2005; Nakata, 2015) has demonstrated that reactivation intervals timed to the point of forgetting lead to longer lasting retention curves. This study replicates that effect at the primary school level: spacing vocabulary reviews over a few weeks resulted in better recall after three months than cramming the same amount of time in one class. The reason for this success is that spacing causes retrieval attempts to occur at moments of partial forgetting, which both strengthens memory consolidation and protects against interference.

Cognitive Explanations for Spaced Retrieval Superiority

There are several cognitive and neuro-educational explanations for the observed superiority of spaced retrieval practice over massed practice. First, according to Pavlik and Anderson's (2005) activation-based model of memory, each retrieval event reactivates and thus consolidates a fading memory trace and makes it more available for future recall. The spaced retrieval schedule allowed for repeated reactivation of vocabulary traces in this manner, which has been found to strengthen memory representation over the long term (Cepeda et al., 2008).

A second explanation is Roediger and Karpicke's (2006) retrieval effort hypothesis. This posits that the effort involved in recalling information improves retention. Recalling something after a short delay requires more effort than remembering something that is still fresh in memory. Therefore, the retrieval attempts of the spaced group were optimally spaced so that recall attempts were neither too easy nor too difficult to produce desirable difficulty (Kornell & Bjork, 2008).

Third, the spacing effect can be explained by the avoidance of cognitive fatigue and attentional saturation. In massed sessions,

repeated testing in short succession results in superficial but not deep encoding due to a lack of cognitive effort (Rawson & Dunlosky, 2011). By contrast, recovery time between distributed learning episodes allows for consolidation in long-term memory (Cepeda et al., 2008), supporting Bahrick and Phelps's (1987) findings on durable retention after years in massed vs. distributed retrieval schedules across days or weeks.

Fourth, spacing may have helped encode contextual variability, meaning that different contexts provide different retrieval paths, as described by Carpenter et al. (2012). It is possible that distributed retrieval allowed the spaced group to learn new vocabulary on different days, in different lessons, and in different emotional and cognitive states than the massed group, who encountered the same material in a single continuous context. Multiple retrieval paths increase the chances for successful recall at a later point.

Finally, Cepeda et al.'s (2008) temporal ridgeline model for distributed practice proposed that a finite review delay (i.e., retention interval) produces an optimal spacing interval (i.e., reactivation interval) that maximizes retention, although the spacing intervals used in this study were likely already quite effective due to variability in encoding and contextual factors. The fixed duration of four weeks, during which all spacing intervals were kept at three 15-minute review sessions per week, likely allowed learners to encounter vocabulary words in slightly different classroom contexts, strengthening the semantic links and cue availability of the new material.

Pedagogical and Classroom Implications

This study has implications for EFL classroom practice, particularly in early language education. It demonstrates that several short learning episodes of 10–15 minutes over a period of several weeks (i.e., spaced retrieval practice) is more effective than one long revision session. Teachers can easily integrate such practice into their daily routines by scheduling short mini-quizzes, oral drills, or flashcard reviews every few days.

As Pashler et al. (2007) argue, one of the most evidence-based recommendations for instruction is to distribute practice over time. The current findings show that even primary school pupils can profit from such organization of lessons even without high-tech learning environments. As spacing is time-neutral (Rohrer & Pashler, 2007), learners would not need to increase their study time, but only to shift it into multiple shorter sessions over an extended period.

In this regard, technological supports may be leveraged. Computer-assisted spaced repetition systems (Nakata, 2011; Nakata, 2015) can automatically schedule adaptive retrieval attempts at ever-increasing intervals, providing individualized and automated feedback while also reducing the teacher's workload. Such programs can be readily implemented into classroom practice or after-school homework routines to maintain students' motivation over time. Simple platforms like Anki or Quizlet could be used to implement the reactivation intervals that have been shown to be effective in the present study.

Another implication of this study is the need for teacher training. Many EFL teachers, especially in primary schools, still place a premium on repetition and present vocabulary through single-session drills and exercises. The cognitive mechanisms of spacing, retrieval practice, and desirable difficulty could be integrated into teacher training and refresher courses to shift vocabulary instruction from receptive exposure to active retrieval. In this regard, teachers can improve learners' metacognitive understanding by explaining to students that it is helpful to study the same words multiple times over a period of weeks or months.

Finally, spacing supports inclusive learning. Shorter learning episodes may be better suited for younger learners with lower attention spans and better match natural cognitive rhythms. Repeated success in retrieving formerly learned vocabulary can also help increase students' confidence and self-efficacy, two factors that are critical to long-term EFL development (Kang, 2016; Dunlosky et al., 2013).

Broader Theoretical Implications

In addition to pedagogical relevance, this study also contributes to theoretical accounts of how temporal distribution affects consolidation and retrieval strength. The results provide ecological validity for the retrieval-based learning framework by showing that lab-observed principles of long-term retention also apply to a real-world language classroom setting.

In this respect, it can be argued that the combination of retrieval and spacing represents a sort of synergistic process, as Karpicke and Blunt (2011) have proposed. Retrieval should be seen as a consolidation enhancer, while spacing can be seen as a way of amplifying the effects of each single retrieval event. In this experiment, it was the interaction of the two that created optimal reactivation conditions and support for the view that retrieval practice is not a testing effect but a central mechanism of durable learning.

The current data also support cognitive theories on "desirable difficulties" (Kornell & Bjork, 2008), meaning that learning is deepened by having learners exert a moderate amount of cognitive effort in retrieval. Spacing has been shown to induce such desirable difficulty by allowing for partial forgetting before recall, thus resulting in deeper encoding. The findings of this study, then, bridge the gap between cognitive psychology and applied linguistics by showing that timing—not just content—affects durability of memory.

Limitations of the Study

While this study has several strengths, it is not without limitations, which are discussed in this section.

Duration Constraint

- The four-week retention window, while valuable for short-term studies, did not account for longer-term retention over several months. Bahrick et al. (1993) showed the spacing effect can last years, and future work could incorporate a delayed follow-up to observe memory decay curves over longer periods.

Sample Size and Generalizability

- The sample size of 60 participants from one private school limits external validity. Factors such as teaching quality, class environment, or individual learner background may vary at other schools, affecting generalizability. Future studies should be replicated at public and rural schools or in multilingual classrooms.

Lack of Technology Integration

- This study employed paper-based recall and classroom quizzes. Integrating technology-assisted retrieval systems, as per Nakata (2011, 2015), could have yielded even more efficient or individualized spacing schedules. Manual and digital retrieval modes could be directly compared to gather richer data on scalability.

Control of Extraneous Variables

- The study equalized instruction time and materials but could

not fully control for motivation, home practice, and parental support, which may have varied between participants and influenced scores. Future research could combine quantitative with qualitative learner feedback to better understand motivational and affective factors in spaced learning.

Measurement Range

- The study's focus was on recognition and recall of 30 target words. For primary pupils, this is a valid measurement range, but future work could expand to include productive vocabulary (speaking and writing) to see if spaced retrieval transfers to active language use.

This study has confirmed that spaced retrieval practice can significantly improve long-term vocabulary retention for primary EFL learners over massed practice. The results support prior evidence from cognitive psychology that distributed learning episodes result in stronger, more durable memory representations (Cepeda et al., 2008; Rohrer & Pashler, 2007; Karpicke & Blunt, 2011).

The cognitive mechanisms that underlie this advantage—reactivation intervals, effortful retrieval, and contextual variability—appear to work in concert to optimize memory consolidation. From a pedagogical perspective, these findings should encourage EFL teachers to implement short and recurring recall activities into their regular lesson plans.

By translating theory into practice, this study can help to bridge the gap between cognitive psychology and EFL teaching. It has shown how a simple adjustment in practice scheduling can yield significant improvements in vocabulary learning and learner confidence.

Case Studies

To extend the classroom findings and illustrate how spaced retrieval can be flexibly implemented across varied learning environments, four representative case studies were developed. Each case demonstrates a distinct instructional condition—traditional, technology-mediated, individualized, and self-regulated—showing how the spacing principle can adapt to different pedagogical contexts while maintaining its cognitive and linguistic benefits.

Case 1: Retention in a low-resource public elementary school

Classroom distribution of recall can be administered in low-resource contexts. In one small-scale study, 30 students aged 9–10 years were exposed to a set of 30 English target words (simple nouns). Short review sessions of the words were distributed over time as part of daily drill activities: the teacher conducted three 10-minute retrieval practice sessions each week, reviewing previously learned vocabulary with picture cues, sentence completion, and peer interrogation.

After 4 weeks, students demonstrated gradual improvement in active recall accuracy, from an initial 59 percent on the first test to 84 percent on the delayed post-test. A group of learners in a control class who had encountered the same vocabulary in a blocked practice condition (practicing all 30 words in one, 30-minute block session at the end of each week) showed only 67 percent retention. In addition to superior scores, students in the spaced condition exhibited greater confidence in pronunciation and a higher frequency of spontaneous word use.

Implications. These results support prior studies in showing that increased frequency of spaced retrieval improves vocabulary consolidation (Cepeda et al., 2006; Carpenter et al., 2012). Teachers reported lower levels of fatigue and better classroom management as well as higher engagement, suggesting that the spacing effect can be achieved through pedagogical consistency in schedule alone without technological support or increased budget.

Case 2: Spacing schedules in a technology-enabled EFL classroom

A second classroom experiment was carried out in an urban EFL program in a digitally equipped elementary school. In this class, students used tablet-based flashcard software that automatically scheduled reviews of the same set of words according to individual recall performance, implementing an expanding-interval schedule described in Nakata (2011) and Pavlik and Anderson (2005).

Participants completed three computer-administered spaced retrieval sessions per week for a duration of 4 weeks, with teachers monitoring progress through a back-end dashboard. Quantitative data collected on software usage indicated that the spaced, technology-assisted group produced a mean score of 87 percent on the delayed-recall test, while a paper-based matched comparison class showed a 78 percent recall level. Qualitative interviews with pupils in the experimental class revealed higher engagement with the activity: students described it as "fun," "interactive," and "easy to follow".

Implications. These findings corroborate the hypothesis that technology facilitates both learner motivation and retrieval efficiency (Nakata, 2015; Karpicke & Blunt, 2011). Automatic adjustment of intervals between trials reduced instructor workload, and self-paced scheduling of vocabulary card reviews also accommodated individual learners' forgetting curves (rate of forgetting). Overall, technology not only improves memory but also supports self-regulated learning behaviors that are essential for long-term language retention.

Case 3: Individual differences and cognitive load

Classroom spacing schedules may need to be adjusted to accommodate individual differences in proficiency. In one quasi-experiment, even within the same class and under the same conditions, learners exhibited large individual differences in attention span, background knowledge, and memory endurance. For strong vocabulary baselines, three spaced retrieval cycles were sufficient to achieve 80 percent recall, while weak learners needed five to six cycles of practice.

Some children showed signs of temporary overload when study intervals were too short or when there were too many unfamiliar items to be learned and retrieved at once. One solution, which was later implemented by teachers, was adaptive pacing of intervals. For low-proficiency learners, teachers introduced shorter sets of only 10 words each before moving on to longer lists, while for advanced learners, they simply increased the spacing intervals.

Implications. These observations are consistent with Rawson and Dunlosky (2011), who stressed the importance of matching an interval to the strength of memory, and with Kang (2016), who recommended more differentiated retrieval schedules. The results suggest that flexible, data-informed spacing intervals can help avoid cognitive overload and ensure equal vocabulary retention across learners of different proficiency levels.

Case 4: Teacher-led versus self-paced retrieval

Teacher-scheduled spacing schedules may affect learner motivation differently than self-paced schedules. In one experiment, two parallel classes completed the same 4-week curriculum, but the

Table 3: Comparative Synthesis of Case Study Findings

Focus Area	*Context*	*Implementation Method*	*Main Outcomes*	*Pedagogical Implications*
Public Elementary School Implementation	Limited-technology classroom	Paper-based distributed recall	84 % delayed retention; higher engagement	Spaced retrieval works without digital tools; low-cost model
Technology-Assisted Spacing	Urban, device-supported EFL program	Tablet flashcards with adaptive intervals	87 % delayed retention; improved motivation	Technology automates review and supports autonomy
Individual Differences & Cognitive Load	Mixed-ability classroom	Adaptive spacing and reduced item sets	Balanced retention across proficiency levels	Personalized intervals prevent overload
Teacher-Led vs Self-Paced Retrieval	Two parallel EFL classes	Structured vs learner-controlled practice	83 % vs 79 % retention; stronger self-efficacy in self-paced learners	Combine scaffolding with gradual independence

groups were otherwise structurally distinct:

- TA Group: Students were told when and in which order to review.
- SP Group: Students practiced their own vocabulary cards, colored-marking the ones they had mastered (green), partially learned (yellow), and found difficult (red), and then selecting themselves which cards to review in each interval.

By the final test, both classes showed high overall performance, with the self-paced group reaching a marginally higher mean score (83 percent vs 79 percent). Interview data suggested that self-paced learners experienced more agency, felt more invested in and responsible for their own progress, while the TA group, in contrast, expressed satisfaction with the structure.

Implications. These findings are in line with Kornell and Bjork (2008), who found that spacing effects are stronger when learners themselves control practice scheduling, as this increased metacognitive monitoring and self-efficacy. However, full autonomy over study schedule may not always be advisable, especially for younger students who may become discouraged by unstructured work; it is likely that a hybrid approach, starting with teacher scaffolding and gradually moving towards self-regulated retrieval, would be the most effective for supporting independent study habits.

General Implications from the Case Studies

In all four cases, we have observed that spaced retrieval performs better than massed practice, independent of the modality, technology, or learners. The technique can be carried out in a simple and systematic classroom routine, as a part of an online platform, or in a highly tailored manner for individual learners.

Digital tools make the method more engaging and automate the scheduling process. However, even the "offline" approach, if applied in a highly regular manner, can produce significant gains. Accounting for differences in learners and providing options between instructor-led and self-guided activities, can help reach an appropriate level of cognitive effort and maximum retention.

Overall, the four case studies not only confirm the theoretical claims of Bahrick et al. (1993), Cepeda et al. (2008), and Dunlosky et al. (2013) that distributing retrieval practice over time can make stronger memory traces, but also allow for highly durable language learning outcomes that can be considered for wider adoption in the EFL context.

Conclusion

The results of this four-week classroom experiment show that spaced retrieval leads to significantly higher EFL vocabulary retention among primary school learners. The performance gains over baseline (delivered over multiple weeks) and the delayed post-test scores of students in the distributed learning group were both higher than those of their massed practice counterparts. These results indicate that information encoded using spaced recall is better consolidated into long-term memory, and suggest that it is more durable and less likely to be forgotten in the long term. This outcome is in line with the evidence from the literature of cognitive and educational psychology reviewed earlier and confirms that the spacing effect and the retrieval practice principle can be extended beyond the learning of word meaning to additional vocabulary knowledge domains, and other primary school learning contexts (Cepeda et al., 2006; Roediger & Karpicke, 2006; Dunlosky et al., 2013).

The observed advantage of spaced retrieval practice supports a key principle of learning and instruction: not only what is being learned but how learning opportunities are distributed matters. When retrieval opportunities are adequately spaced out and repeated over time, they appear to engage learners in deeper processing that more effectively strengthens memory traces and ensures longer-term retention (Karpicke & Blunt, 2011; Kang, 2016). By contrast, the findings suggest that the massed practice group's performance quickly became overlearned in a short time without lasting (long-term) effects, a common but negative pattern often seen in classroom studies of intensive practice (Rohrer & Pashler, 2007).

In terms of immediate classroom application, the study results have clear implications for teaching and learning design and lesson planning in primary EFL settings. Instructors should be encouraged to reduce or avoid intensive cramming sessions in favor of shorter, spaced review activities distributed over time that require learners to actively recall previously studied information. These distributed learning opportunities could easily be incorporated into regular lessons and would not have to lengthen school hours. Instead, for example, they could be a matter of 10–15 minutes of review either at the start or the end of each week. In practice, teachers may easily turn to low-stakes quizzes, games, or digital flashcards that make students produce meanings instead of passively rereading and copying from lists. In addition to improved retention, such approaches can also boost student engagement, motivation, and self-efficacy in language learning.

In this regard, the study also points to the role of technological spacing aids. Computer-assisted learning systems and smartphone applications that schedule repeated review opportunities based on learner performance or the forgetting curve (e.g., Anki or Quizlet) can make spaced retrieval even in large classes practical (Nakata,

2011; Pavlik & Anderson, 2005). Embedding these types of tools into primary EFL curricula could allow for extending the effects of the current small-scale experiment to larger and more diverse student populations.

This experiment has proven the cognitive benefits of distributed learning over massed practice but is limited by its duration, sample size, and single-school setting. Therefore, future research should address the following aspects and questions. First, studies with larger samples and longer intervals are needed to investigate the effects of spacing over longer terms. How, for example, does vocabulary retention change if students are to be tested in an end-of-term delayed post-test three or six months later? Second, how do the findings differ for other learner profiles and age groups? Are developmental or motivational factors associated with spacing interval effects on selective attention and working memory? Third, how do teacher-mediated and technology-supported retrieval schedules compare? In what circumstances do teacher-controlled or self-paced digital applications lead to the most sustainable long-term outcomes, given variations in available resources?

To sum up, the study provides evidence-based support for the application of spaced retrieval in EFL instruction at primary schools. The research shows that spaced retrieval practice is more effective for EFL vocabulary retention among primary school children than its massed practice alternative. The advantages of distributed learning and repeated testing clearly demonstrated in the results suggest that in addition to having multiple opportunities to learn a word, primary EFL students need well-timed repeated opportunities to recall information over time. This, in turn, can be easily applied to actual instruction through the reduction or complete avoidance of intensive cramming and the introduction of distributed, repeated retrieval practices over several weeks. As such, it can potentially lead to more durable vocabulary learning, better language learner autonomy, and higher EFL proficiency in general among young learners.

References

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological bulletin, 132(3), 354.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological science, 17(3), 249-255.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. Science, 331(6018), 772-775.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public interest, 14(1), 4-58.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. Psychological science, 19(11), 1095-1102.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & 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.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. Psychology and aging, 25(2), 498.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. National Center for Education Research.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. Psychological Science, 4(5), 316-321.

Pavlik Jr, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. Cognitive science, 29(4), 559-586.

Nation, I. S., & Nation, I. S. P. (2001). Learning vocabulary in another language (Vol. 10, pp. 126-132). Cambridge: Cambridge university press.

Schmitt, N. (2008). Instructed second language vocabulary learning. Language teaching research, 12(3), 329-363.

Bahrick, H. P., & Phelphs, E. (1987). Retention of Spanish vocabulary over 8 years. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(2), 344.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. Computer Assisted Language Learning, 24(1), 17-38.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. Psychological science, 19(6), 585-592.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough?. Journal of Experimental Psychology: General, 140(3), 283.

Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. Policy Insights from the Behavioral and Brain Sciences, 3(1), 12-19.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning?. Studies in Second Language Acquisition, 37(4), 677-711.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. Current directions in psychological science, 16(4), 183-186.

Logan, J. M., & Balota, D. A. (2008). Expanded versus equal interval spaced retrieval practice. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(5), 1358–1369.