

# UNIVERSIDAD DE ALMERÍA



MÁSTER EN PROFESORADO DE EDUCACIÓN SECUNDARIA OBLIGATORIA Y  
BACHILLERATO, FORMACIÓN PROFESIONAL Y ENSEÑANZA DE IDIOMAS

ESPECIALIDAD EN LENGUA INGLESA

**Curso Académico:** 2015/2016

**Convocatoria:** Junio

**Trabajo Fin de Máster:** Spaced Retrieval Practice Applied to Vocabulary  
Learning in Secondary Education

**Autor:** Héctor Daniel León Romero

**Tutora:** Susana Nicolás Román

## **ABSTRACT**

Spaced retrieval practice is a learning technique which has been long studied (Ebbinghaus, 1885/1913; Gates, 1917) and long forgotten at the same time in education. It is based on the spacing and the testing effects. In recent reviews, spacing and retrieving practices have been highly recommended as there is ample evidence of their long-term retention benefits, even in educational contexts (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). An experiment in a real secondary education classroom was conducted in order to show spaced retrieval practice effects in retention and student's motivation. Results confirm the evidence, spaced retrieval practice showed higher long-term retention (26 days since first study session) of English vocabulary words compared to massed practice. Also, student's motivation remained high at the end of the experiment. There is enough evidence to suggest educational institutions should promote the use of spaced retrieval practice in classrooms.

## **RESUMEN**

La recuperación espaciada es una técnica de aprendizaje que se lleva estudiando desde hace muchos años (Ebbinghaus, 1885/1913; Gates, 1917) y que al mismo tiempo ha permanecido como una gran olvidada en los sistemas educativos. Se basa en los efectos que producen el repaso espaciado y el uso de test. En recientes revisiones de la literatura se promueve encarecidamente el uso de estas prácticas, ya que aumentan la retención de recuerdos en la memoria a largo plazo, incluso en contextos educativos (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Se ha llevado a cabo un experimento en una clase de E.S.O. con el objetivo de mostrar los efectos de la recuperación espaciada en la retención de recuerdos y en la motivación de los alumnos. Los resultados confirmaron lo esperado, la recuperación espaciada mostró una mejor retención a largo plazo (26 días desde la primera sesión de estudio) de vocabulario en inglés en comparación con el estudio concentrado de contenidos. A su vez, la motivación de los estudiantes se mantuvo alta una vez finalizado el experimento. Hay suficiente evidencia para sugerir que las instituciones educativas deberían promover el uso de la recuperación espaciada en las aulas.

## TABLE OF CONTENTS

1. Introduction: justification, objectives and methodology .....	3
2. Theoretical Framework .....	5
2.1. How memory works .....	5
2.2. The spacing effect.....	6
2.3. The lag effect .....	8
2.4. Cognitive theories on distributed practice .....	10
2.5. The testing effect.....	12
2.6. Cognitive theories on the testing effect.....	15
2.7. Spaced retrieval practice .....	16
2.8. Spaced repetition systems.....	17
3. Spaced retrieval practice in teaching .....	19
3.1. Introduction.....	19
3.2. Implementation in the classroom .....	21
a. Objectives .....	21
b. Context .....	21
c. Participants .....	21
d. Methodology .....	21
e. Results and discussion .....	23
f. Limitations of the experiment.....	26
4. Conclusions.....	27
References.....	30
List of figures and tables.....	35
Appendix A: Study sessions material.....	36
Appendix B: Study tests (with answers).....	37
Appendix C: Satisfaction survey .....	38

# SPACED RETRIEVAL PRACTICE APPLIED TO VOCABULARY LEARNING IN SECONDARY EDUCATION

## EL USO DE LA RECUPERACIÓN ESPACIADA APLICADA AL APRENDIZAJE DE VOCABULARIO EN LA ENSEÑANZA SECUNDARIA

Héctor Daniel León Romero

### 1. Introduction: justification, objectives and methodology

Nowadays, the Spanish educational system is lacking of empirical methods for learning retention in students. There is an established belief that the significance of teaching is delivering content (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Once content can be partially elicited, educators tend to assume that it has been learnt and that any further rehearsal is not needed; thus, being able to move to other material to repeat the same method (Karpicke & Roediger, 2008). It seems like temporal distribution of learning materials is not being taken into account; one exam per term seems to be the norm and intensive courses like summer camps or immersion courses appear to be fashionable (Pashler, Rohrer, Cepeda, & Carpenter, 2007).

As students pass courses, learning changes from being largely supervised to being rarely supervised. Students are expected to self-control their own education in higher levels but this lack of regulation usually makes demotivated students mass their practice around the exam instead of spacing it. According to Kang (2016), grade point average (GPA) seems to be higher in students who space their learning. Furthermore, Kang states that although college students seem to be conscious about the benefits of spaced practice (in contrast with cramming), they affirm to regularly recur to cramming before tests. Massing results in low retention of information, even in the short-term,

leading to frustration and dramatic career deviations (Hopkins, Lyle, Hieb, & Ralston, 2015).

Recently, the poor results on retention in education have started a large interest in cognitive psychologists to find the best techniques to promote superior retention (Hopkins et al., 2015). Dunlosky et al. (2013) reviewed the literature on 10 widely used learning techniques and gave typically used techniques as summarization, highlighting, and rereading a low utility assessment. Only two techniques received the highest utility assessments: practice testing and distributed practice. This review showed large evidence of their benefits to learners even in educational contexts.

Distributed practice was first studied in the late 1800s (Ebbinghaus, 1885/1913) and practice testing in the early 1900s (Gates, 1917)<sup>1</sup> and their effectiveness – even in educational contexts (see Spitzer, 1939) – was already proven then. In modern days, the case of testing practice is particularly noteworthy given that there is an ongoing trend in education which goes against testing in classroom instruction (Roediger & Karpicke, 2006a). Nonetheless, if the evidence is so clear, why it is not implemented in the classrooms?

These are the **objectives** of this research article:

- To review the literature on spaced retrieval practice, its cognitive mechanisms, and the different ways to implement it, with a special emphasis on education.
- To measure spaced retrieval effects in a real Spanish Compulsory Secondary Education classroom.
- To find the effects spaced retrieval practice can have in student's motivation.

In order to meet the objectives, literature on the topic has been reviewed and an experiment has been conducted in the classroom. The followed **methodology** for the review of the literature has been analytical and not merely descriptive. The terms “spaced retrieval,” “testing effect,” “spacing effect,” “distributed practice,” “cognitive

---

<sup>1</sup> There are even prior antecedents in which the testing effect is mentioned (Bacon, 1620/2000; James, 1890)

psychology,” “spaced repetition systems,” and “learning techniques” have been looked for in the literature in depth, mostly via Google Scholar (2016). The **methodology** used for the experiment is based on Cepeda, Pashler, Vul, Wixted, and Rohrer (2006) and Pashler, Rohrer, et al. (2007) but it has been adapted to the context. Contents were taught during a first study session; reviewed and retested in following sessions (interstudy interval) and retested again (retention interval) 26 days after first study session. Then, a satisfaction survey was provided to check student’s motivation.

## 2. Theoretical Framework

### 2.1. How memory works

In education, the goal is to enhance long-term memory but, what types of memory do we have and how is the process to reach long-term retention? Human memory is a complex mechanism that is used to encode, store, and retrieve information (Davey, 2014). It is thought to have a series of independent systems which are in charge of different sorts of memory. This idea has evolved from neurodegenerative diseases and damaged nervous systems, which impair different types of memory (Davey, 2014). The current model is based on the Atkinson-Shiffrin model (Atkinson & Shiffrin, 1968), or multi-store model, which made a division of three separate components: sensory, short-term, and long-term memories. Since then, there have been some changes in the theory; short-term memory is now called working memory (Benigas, 2015) and each of these three basic components is divided into some other processes or components (Baddeley, 2000).

According to Santrock (2011), information received via our senses (sight, hearing, touch, taste and smell) is treated in the sensory memory. This memory lasts from a fraction of a second to a few seconds so it is highly necessary for a learner to get the relevant information before it fades. Information that has been paid attention to, goes to the working memory, which looks like a working table instead of a memory storage.<sup>2</sup>

---

<sup>2</sup> Short-term memory is still a widely used concept and theory neutral to show talk about retention. It is supposed to last around 30 seconds and it has span limitations (e.g., in a classic test, series of digits and their orders have to be recalled and the frequent number of items recalled is about 7) (Davey, 2013).

Here, information is actively manipulated and wired in order to solve problems, take decisions, and understand oral and written language. Finally, long-term memory can possibly retain limitless information for the rest of a person's life. Some proven mechanisms to help to consolidate information into lasting memories are these: meaningful associations, emotional influence, or a good sleep.<sup>3</sup> The way information is forgotten is not clear but according to Davey (2013), there are two hypotheses: information degrades or interferences appear.

## 2.2. The spacing effect

This section is going to focus on the *spacing effect*, referring to the superior learning effect of spaced episodes in comparison with massed study. As the concepts can be confusing in this and the following sections, the same terms as Cepeda, Pashler, et al. (2006) are going to be explained. Cepeda, Pashler, et al. (2006) performed a meta-analysis involving more than 14,000 participants, one of the largest reviews in the field.

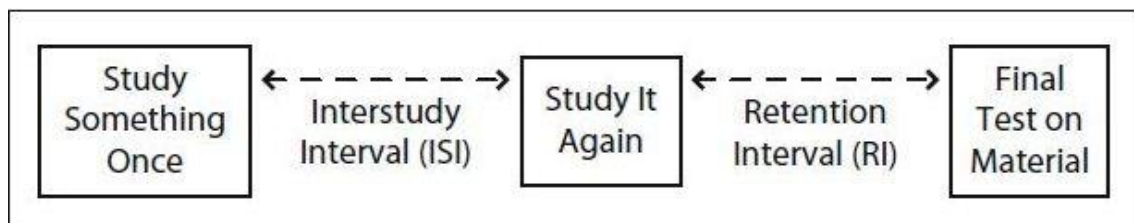


Figure 1. The basic scheme of a spacing experiment. Two opportunities are given to learn the same learning material. ISI and RI varies from one study to another. Source: adapted from "Enhancing learning and retarding forgetting: Choices and consequences," by H. Pashler, D. Rohrer, N. J. Cepeda, and S. K. Carpenter, 2007. *Psychonomic bulletin & review*, 14(2), p. 188.

In a classic spacing study (Figure 1), the information is studied at least two times and then tested. The interval between the first and the last study is called *interstudy interval (ISI)* and the interval between the last study and the final test is called *retention interval (RI)*. In these studies, *massed* learning should consist in single items presented

---

<sup>3</sup> This is a brief and simplified explanation of the contemporary knowledge on the topic.

for twice the time a spaced item would, and with less than 1 second between each item.<sup>4</sup> *Spaced or distributed* learning must have separated session studies of the same item of 1 s or longer. Thus, the term *spacing effect*, as said in the previous paragraph, is the superior learning effect during spaced episodes in comparison with massed study; while the term *lag effect* refers to different ISIs and RIs between different studies. The term *distributed practice* (some other forms are *spaced repetition* or *spaced practice*) comprehends the spacing and lag effects, so it refers to the general spacing of learning elements.

Once the terms are clear, data is going to be reviewed. Spacing has been being studied for more than a century (Ebbinghaus, 1885/1913) so there is vast evidence of it. In the previously mentioned meta-analysis, Cepeda et al. (2006) found that, on average, spaced practice produced a 15% of improvement in results compared with massed study, regardless of retention interval; it appears to apply to people of all ages, from children to adults. Additionally, every study surveyed with a RI longer than 1 month showed a benefit of studying across weeks or months, instead of massing the study in 1 day.

In terms of types of knowledge, the spacing effect is present in learning facts, lists, concepts, complex theoretical concepts, skill and motor learning and in-class science, vocabulary, foreign vocabulary, maths, history, statistics, spelling, and many more (Balota, Duchek, & Logan, 2007; Dunlosky et al., 2013; McDaniel, Fadler, & Pashler, 2013; Smolen, Zhang, & Byrne, 2016; Sobel, Cepeda, & Kapler, 2011). Moreover, there is evidence of spacing effect in molluscs, flies, bees, rodents and non-human primates (Smolen et al., 2016).

To conclude, the spacing effect is a reality and it should be taken into account in our educational system by repetitions of the material, not only during the same academic year, but also across courses so as to stabilize the learnt knowledge.

---

<sup>4</sup> In a massed study, the items cannot be presented in an ABCABC sequence because that would entail a type of spacing called *interleaving*; instead, the sequence should have an AABBC form, without lapses between the same or the other materials (Kang, 2016).



### 2.3. The lag effect

When scheduling the ISI and the RI,<sup>5</sup> infinite possible variations are found and some questions can be raised: is there an optimal spaced interval? If some material is taught today, when should it be reviewed? (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012) All these possible variations of time and amount of reviews are encompassed in the so-called *lag effect*.

Distributed practice studies require two or more learning episodes. When the ISI presents more than two learning episodes, they can be *expanding* (e.g., 1-4-10), *equally spaced* (e.g., 5-5-5), or *contracting*<sup>6</sup> (e.g., 10-4-1) (Cepeda et al., 2006). When studying the different types of intervals, not only spacing matters but also the study method. The rate of forgetting is highly influenced by the type of learning method used; thus, repeated testing in a given interval seems to slow forgetting when compared with no testing or having just 1 testing episode in the ISI (Karpicke & Roediger, 2008). This is due to the *testing effect*, which will be discussed in a following section.

In a large study on the lag effect of two learning episodes, Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) studied different ISIs and RIs. Taking 1,350 individuals in 26 different interval variations, with ISIs ranging from 0 to 105 days and RIs ranging from 7 to 350 days, to study recognition and recollection of trivia facts. They found that as the RI increased, the optimal ISI also increased; the optimal ISI being about 10–20% of the RI (Figure 2). Therefore, the most efficient approach to study for a test 1 week away would be a 1-day ISI; for a test 1 year away, an ISI of about 2 months; and for a 5-year away test, an ISI of 6 to 12 months would be ideal. Pashler, Rohrer, et al. (2007) also found the same 10–20% optimum ISI/RI percentage in foreign vocabulary learning. This ratio can have many applications, among them, the studying of core content in the educational system across several course years (Dunlosky et al., 2013). It is not obligatory to stick to the optimal rate but it is advisable since overlearning is inefficient from a long-term perspective (Pashler, Rohrer et al., 2007).

---

<sup>5</sup> See Figure 1 and pages 6 and 7 for an explanation of these acronyms.

<sup>6</sup> Contracting learning intervals seem to be inefficient and they are not given much attention in the literature.

The previous paragraph discussed ISIs of only two learning episodes but there is also an enduring debate in the field regarding three or more learning episodes: what is better, expanding or equally spaced intervals? Reviews tend to find similar levels of learning in long-term retention (Balota et al., 2007) or insufficient data (Carpenter et al., 2012; Cepeda et al., 2006).

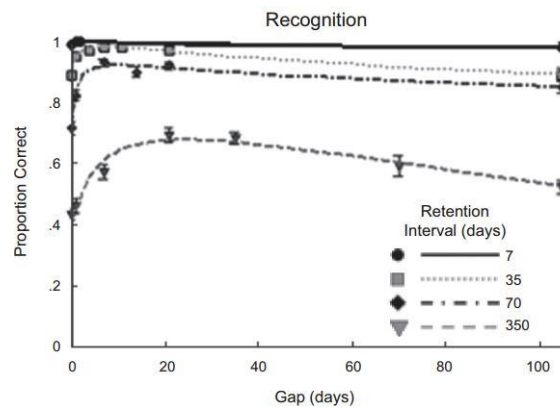


Figure 2. Performance on recognition of trivia facts. For every given RI, there are several ISIs (or gap, in the figure). Source: adapted from “Spacing effects in learning a temporal ridgeline of optimal retention,” N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, 2008.. *Psychological science*, 19(11), p. 1098.

Karpicke & Roediger (2007), found better long-term retention for equally spaced retrieval in comparison with expanding retrieval, and better short-term retention for expanding retrieval.<sup>7</sup> These results led to the conclusion that equally spaced intervals created a “desirable difficulty” provided that the first test is done later than in an expanding interval practice.

There is an extended belief that expanding practice is better than equally spaced practice. Some classic works like Landauer & Bjork’s (1978), showed benefits for expanding practice and, similarly, some recent research also does (Kang, Lindsey, Mozer, & Pashler, 2014). In fact, It seems natural to believe that the best moment to remember any information is when it is about to be forgotten.<sup>8</sup> The Pimsleur method (Pimsleur, 1967), the Leitner system (Leitner, 1986), or the SuperMemo algorithm (Wozniak, 1990)

<sup>7</sup> Every conducted experiment had around 50 undergraduate participants.

<sup>8</sup> Landauer & Bjork (1978) had 468 participants but Kang et al., (2014) limited their research to 37 subjects (in the latter study, participants’ age ranged from 20 to 63 years).

are all learning methods based in this hypothesis. Leitner, for instance, based his work principally on the antique studies of Ebbinghaus (1885/1913). Nowadays, many software systems as Anki (Ankirs, 2016), Mnemosine (Mnemosyne-proj, 2016), or Mosalingua (MosaLingua, 2016) are grounded on Wozniak's algorithm.<sup>9</sup>

Expanding retrieval is considered to improve memory in memory-impaired populations due to the no-error learning it promotes (Karpicke & Roediger, 2007). The functioning mode of this method is supposed to work this way: a retrieval short after the last study session guarantees high retention by slowing forgetting, and thus, subsequent tests are progressively expanded so as to continue to be effortful without forgetting (Kang, 2016)

Overall, it appears that there is no need to overemphasize the significance of the expanding vs. fixed spaced practice dualism. It is possible that the inconsistent data from the literature can be a result of differences in difficulty, types of review, or time frame (Kang, 2016). Further research may show some answers as in the ideal two-learning-session interval. This optimal ISI from 10% to 20% should be planned and discussed with the students to anticipate future needs; thus, not only the students would be partly in charge of their own learning but also, retention would be higher, and they would feel more motivated for that. Likewise, core contents would have to be chosen by the teacher, cleaning the curricula from clutter, and making it clearer for the students. Anyway, long-term speaking, any spacing form would promote higher retention in comparison with massed schedules (Carpenter et al., 2012). Efforts should first be put on this single enhancement given that its only implementation would already be a difficult step and at the same time a great success.

#### 2.4. Cognitive theories on distributed practice

Distributed practice combines spacing and lag effects. The mechanisms behind these techniques are not well understood; nevertheless, several theories have been

---

<sup>9</sup> Spaced repetition systems will be discussed in a following section.

proposed. The three most well-known hypothesis to explain distributed practice are deficient-processing theory, encoding variability theory, and study-phase theory.

According to Smolen et al. (2016), **deficient-processing theory** suggests that in massed training some necessary routes to create memories are not successfully activated. There are some variants of deficient-processing theory; for instance, *consolidation theory*, which was first applied to short term memory but that has been adapted to long intervals. Thus, consolidation variance is based on these two assumptions: the neural circuit state after a first exposure to the learning material has some peculiarities that inhibit the consolidation of memory traces from a second exposure right after; and that the likelihood that the second exposure to the learning material effectively strengthens the first exposure decays with time.

**Encoding variability theory** supports that memory traces include elements of the learning context, and spacing apart learning trials usually entails different contexts; thus, the more contexts, the more associations can be made, and the more robust a memory will be (Balota et al., 2007). According to this theory, equally spaced practice would yield higher results than expanding practice in the long term. This is because expanding practice usually has a prior exposure to the material, so the first two contexts would be rather similar (Smolen et al., 2007).

**Study-phase retrieval theory** posits that each spaced learning episode causes reactivation of a memory trace, while in massed learning, supposedly, memory traces cannot be reactivated because they are still being used (Smolen et al., 2007). Furthermore, long intervals are assumed to produce a decline in the memory trace, making it irretrievable.

Encoding variability, study-phase retrieval, and the explained consolidation theory seem to be in line with Cepeda et al. (2008) finding of an optimal gap between study sessions. Some other deficient-processing theory variances refer only to working memory, which lasts minutes or even seconds; however, optimal gaps are usually found in larger intervals so these versions are dismissed by Cepeda et al. (2008).

Some other theories have been proposed, like the desirable difficulty theory,<sup>10</sup> but in general, distributed practice mechanisms are not well understood and further evidence is needed. Study of the brain via magnetic resonance imaging (MRI) is evolving swiftly, and could provide some answers on the distributed practice cognitive devices.

## 2.5. The testing effect

The benefits of testing, not only as a measuring tool but as a proper learning device have long been forgotten in our education system and, in general, outside of cognitive psychology (Roediger & Karpicke, 2006a, 2006b). Tests' poor reputation may possibly come from different reasons: the fact that grading can influence student's self-opinion, subsequently lowering self-expectancy, self-motivation, and latter results; their latent association with measure, control, and the consequent stress; or the cognitive higher efforts it requires as compared to extended methods like rereading, highlighting, or summarizing. Neither students seem to like tests nor do teachers like giving them, so test avoidance should be expected in educational settings.

Testing effects need clarification on the issue that they can bring direct or indirect benefits. On *indirect* (or mediated) effects, Roediger and Karpicke (2006a) provide some illuminating examples: given that students usually study for tests, if testing is used in classrooms on a regular basis, the amount of time spent learning will also be higher and it will be more spaced;<sup>11</sup> if the students receive proper feedback after a test, the study time be higher as a result; and frequent testing could also influence the way student's learn on their own. On the other hand, the so-called *testing effect*<sup>12</sup> (the direct effect) refers to the improvement of long-term retention after an initial study of the learning material, compared to not taking a test or to restudying the content (Hopkins et al., 2015; Van Gog & Sweller, 2015).

As it has been stated in the introduction of this document, the testing effect has been studied for over a century (Gates, 1917; Spitzer, 1939). In 1939, Spitzer draw a

---

<sup>10</sup> According to Karpicke & Roediger (2007), a greater recalling effort involves better retention; thus, a desirable difficulty implies high efforts to be retrieved.

<sup>11</sup> See page 7.

<sup>12</sup> Also called the *retrieval practice effect*.

series of conclusions from a large study of the testing effect with 3,605 sixth-grade pupils. Spitzer recommended the use of testing in elementary schools due to its proven effectiveness in aiding retention, and also its usefulness in fixing erroneous knowledge when tests were corrected by either the teacher or the pupils themselves.

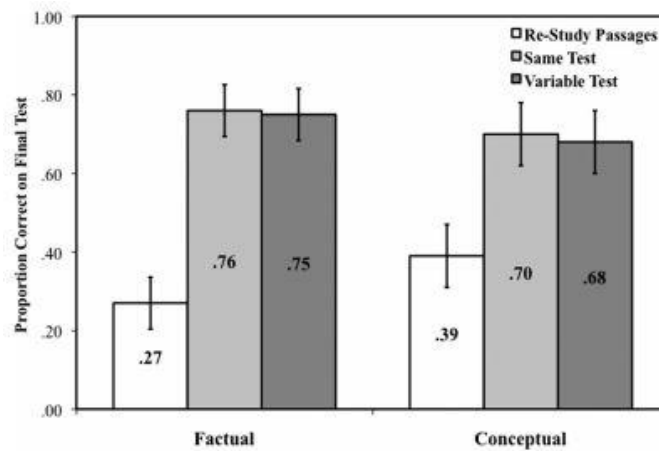


Figure 3. Six passages were either studied or tested and in the same session either restudied (Re-Study Passages), or retested (Same Test and Variable Test). The graphic shows proportion of correct responses after a 1-week interval. Source: adapted from "Repeated testing produces superior transfer of learning relative to repeated studying," by A. C. Butler, 2010. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), p. 1123.

Classroom studies cannot be as controlled as laboratory ones so in-class evidence have to be carefully examined. This leads to the question of whether there is enough evidence to affirm that testing provides better results than restudying in classroom settings. Recent meta-analyses (Dunlosky et al., 2013; Roediger & Karpicke, 2006a; Rowland, 2014) on the topic and contemporary studies (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013) all point to the same conclusions: testing is highly effective even in formal education.

Testing appears to enhance memory in peculiar ways. For instance, it can slow forgetting as compared to restudying a given material (Carpenter, Pashler, Wixted, & Vul, 2008; Karpicke & Roediger, 2008). What is more, testing effect persists in the long term even when feedback is not given (Karpicke & Roediger, 2007). In Roediger and Karpicke (2006b), although repeated-studying showed better results in a 5-minute-delay test after first study, in the consequent 2-day and 1-week-delay tests, the repeated-

testing group outperformed the repeated-study one. In the 1-week-delay test, testing beat studying by a 21% even with the former receiving no feedback and the latter being exposed to the learning material four-fold (3.4 times vs. 14.2 times).

In another study, this one on vocabulary learning (Swahili), Karpicke and Roediger (2008) studied four different groups with different study-test schedules until words were learnt. The two groups which had to test every word every testing time, correctly recalled 80% of the words after a week, in contrast, the two groups which only had to recall every testing time the words they could not retrieve in the previous test, correctly recalled only about 36% (less than the half of the other group). Conclusions can be drawn from these experiments. Retrieval per se is a powerful mechanism and the use of it should be implemented more regularly in the schools. Long-term retention can be increased if retrieval methods are used instead of recognition ones like multiple choice tests (see also Kang, McDermott, & Roediger, 2007).

The cognitive mechanisms of the mentioned effect are not well understood yet but light is being shed on this issue progressively. For example, it is known that the repeated-study groups show more confidence in their recalling ability than the repeated-testing groups (Roediger & Karpicke, 2006b); this can indicate why testing is not an extended self-study method. Additionally, errors associated with guessing in testing seem to not cause any problem in retention provided that feedback is ultimately provided (Pashler et al., 2007). Finally, in cued recall vocabulary tests, it has been showed that after a study session of word pairs (A–B), the form of retrieval do not need to be the same in consequent tests with feedback (A–? vs. ?–B) to beat repeated-study groups (Karpicke & Roediger, 2008) and that recalling in one direction (A–?) facilitates recalling in the opposite direction (?–B) (Pashler, Rohrer et al., 2007).

Vocabulary learning has been well studied as for testing and a typical vocabulary study method implies testing: flashcards. Nevertheless, flashcards can be used to learn anything. However, Van Gog and Sweller (2015) found that if the tested material were complex, testing effect was not an improvement. Despite the fact this study was soon rebutted (Karpicke & Aue, 2015), testing of complex materials need to be further studied.

In conclusion, testing has been showed to be one of the most effective study methods with proven evidence also in educational settings (Dunlosky et al., 2013). Maybe because of the testing effect being slightly counterintuitive, it is not well known in education; that is why efforts by the institutions must be put to spread its benefits even on a daily basis.

## 2.6. Cognitive theories on the testing effect

Although testing is often regarded as a measuring technique, it entails great long-term retention benefits – as it can be seen in the previous section; but how does the testing effect work? Three different approaches will be discussed in this section: the elaborative retrieval theory, the transfer appropriate processing theory, and the unspecific-goal perspective.

**Elaborative retrieval theory** (Carpenter, 2009) binds together several theories. It is based on the concept of spreading activation and the degree of semantic elaboration, which are normally connected with encoding and retrieval processes (Wing, Marsh, & Cabezaa, 2013). *Spreading activation* refers to the strengthening of existing retrieval cues and the formation of new ones when retrieving; even if the targeted memory is not recalled, related elements are reinforced. Spreading activation has been upheld on the fact that retrieval yields better results than recognition (Kang et al., 2007). On the other hand, the *degree of semantic elaboration* is subject to the spent mental effort dedicated; thus, weak concept relations lead to better results than strongly related concepts when testing due to the higher effort needed to relate the weakly related concepts (Wing et al., 2013).

**Transfer appropriate processing theory** (Morris, Bransford, & Franks, 1977) posits that a test will yield better results for previously tested groups than for restudy groups because the measure method is the same for the repeatedly tested groups. The third theory is the **unspecific-goal perspective**, which upholds that unspecific problems, such as free recall problems, may foster higher results than specific problems, such as cued recall problems (Endres & Renkl, 2015).



The testing effect has proven its benefits compared to most study techniques (Dunlosky et al., 2013), however, as in distributed practice, the cognitive mechanisms behind the testing effect are not clear. Although elaborative retrieval theory is gaining popularity (Endres & Renkl, 2015), further evidence is required.

## 2.7. Spaced retrieval practice

Benefits of spacing and testing in learning have already been discussed in previous sections of this document. It is logical to believe that mixing both techniques will have greater benefits than using none or just one of them; the method which brings them together is called *spaced retrieval* (Hopkins et al., 2015).

Carpenter, Pashler, and Cepeda (2009) studied retention of history facts in 8<sup>th</sup> grade students. Students reviewed the material either 1 week after the initial learning practice (immediate review), 16 weeks after (delayed review), or it was not reviewed at all (massed learning). In a final test 9 months after the first study practice, spaced study yielded higher results than massed study but lower results than spaced retrieval with feedback. The delayed-review group performed significantly better than the immediate-review one when restudying and retesting, but not for the massed learning group (for similar results in vocabulary learning, see Bloom & Shuell, 1981). According to these results, for the long-term retention, spaced restudying yields better results than massing but spaced retrieval outperforms both of them.

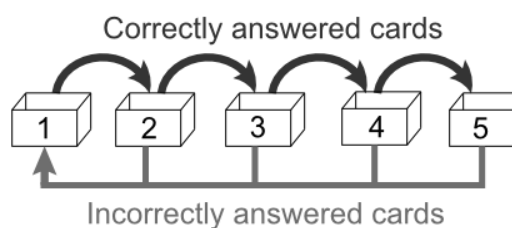
It is notorious that although spaced retrieval is not known in educational contexts, it has been largely been used in other contexts such as memory-impaired population treatments – e.g., older adults or people with dementia – (Karpicke & Roediger, 2007). They use an altered variation of the technique which is called *spaced retrieval training* (SRT). SRT has been developed over the last 27 years and it is based on a reduction of the learning material and the use of extended intervals. The number of clinicians using SRT has increased (Benigas, 2015), while spaced retrieval in education is still at a standstill.

Combining the two study techniques with the highest rated utility in recent reviews of the literature (Dunlosky et al., 2013; see also Pashler, Bain et al., 2007), yields

even greater results than using just one of them or not using them at all (Hopkins et al., 2015; Kang, 2016).

## 2.8. Spaced repetition systems

Evidence on distributed practice have inspired the creation of some distinctive learning methods. The Pimsleur method (Pimsleur, 1967) was one of the first to be developed. It is used to learn languages via audio materials and extending spaced repetition. Nowadays, even though Paul Pimsleur died in 1976, the Pimsleur method is still being used; recently, it has been used by the military U.S. troops to learn Pashto, a largely spoken language in Easter Afghanistan and Northern Pakistan (Pimsleur.com, 2016).



*Figure 4.* The Leitner system. Souce: Adapted from Leitner system. Retrieved June 16, 2016, from [https://en.wikipedia.org/wiki/Leitner\\_system](https://en.wikipedia.org/wiki/Leitner_system).

A different method is the Leitner system (see Figure 4; Leitner, 1986) which is based in the studies of Ebbinghaus (1885/1913). It was originally devised in 1970 and introduced flashcards for the first time (Kuhl & Marolt, 2015). The Leitner system is also based in expanding intervals. It requires several boxes, where flashcards are to be put. New flashcards are placed in a first box and if they are successfully recalled, they are moved to the second box and so on. If they cannot be recalled, they have to be moved to the first box. The first box is checked every day, the second box every two days, and so forth to the total of 5 boxes. When a flashcard is positively recalled in the 5<sup>th</sup> box, it is supposed to be learnt and not to be repeated again.

In 1987, Wozniak wrote the SuperMemo algorithm to use expanding spaced repetition of flashcards in computers (Wozniak, 1990). This was a valuable precedent as, for the first time, human decisions on when to repeat a learning item were not required. Once the user has tried to retrieve a flashcard, several options of difficulty are showed and the learner has to decide the degree of easiness or difficulty. The easier a flashcard, the more extended the following spaced interval and vice versa (Wozniak & Gorzelanczyk, 1994).

Spaced repetition systems (SRS) have multiplied since then but most of them keep being either based on the Leitner system or on the SuperMemo algorithm. The most famous SRS software at the moment is Anki, which is based in the SuperMemo algorithm (Ankisrs, 2016). Anki is a free software which uses flashcards; it is highly customizable and user friendly. Flashcards can be created by the user or imported from other sites or users. Default settings involve expanding intervals but they can be easily modified to equally spaced intervals or to any desired lag. Flashcard creation is limitless but to name a few characteristics, images, videos, and audios can be attached by drag and drop, cards can be asked in both ways (A-B vs. B-A), and cloze exercises can be created without much effort. Furthermore, cards can be synchronized across different platforms (e.g. Windows, OS X, Android, iOS, web browsers, etc.)

The benefits of electronic flashcards do not reside only in the automatic scheduling of cards, they often imply testing, so they are great devices to practice spaced retrieval. These are some guidelines on flashcards creation: questions should require short answers in order to be easily stored; short answers promote higher retention than multiple choice recognition quizzes (Kang et al., 2007); and also, the use of pictures should be adapted according to objectives, as they has been showed to produce overconfidence, and consequently, lower retention (Carpenter & Olson, 2012).

These SRSs can be rather useful when properly used, nevertheless, adapting them to educational contexts can be complicated. Firstly, learners are dependent on technologies which will not be affordable to some families. Secondly, it is difficult to measure the correct use of any SRS because they generally depend on personal appreciations on learning (e.g. hard, good, or easy). Lastly, students need to be

motivated to use them independently; low usage has been shown in studies (Pérez Chamorro, García Álvarez de Perea, Casasola Balsells, 2015).

SRSs can be potent mechanisms to automatize spacing and enhance learning. At the moment, SRSs are not very popular, but some flashcards systems are (e.g. Cram.com and Quizlet.com) and they are starting to introduce spacing modes. Gaining popularity seems just a matter of time. On the other hand, currently, SRSs have been shown difficulties adapting to our educational system. Further social and software developments have to emerge in order to extend their usage.

## 3. Spaced retrieval practice in teaching

### 3.1. Introduction

As it has been showed in previous sections, evidence on the spacing and the testing effects has long been proven in educational contexts; the problem being that teachers who base their teaching methods on actual science are not common. According to Kang (2016), widespread techniques are trusting on familiar methods – like the ones they were taught with – or depending on the teacher’s intuition. In addition, conservative instruction, which tends to favour massed learning, seems to be the norm. Besides, students’ discipline to self-study using science-based methods is rarely consistent (Karpicke & Roediger, 2008). Some instructions are needed in order to guide teachers and influence learners on the best strategies they could use to foster retention.

Hypothetically, one of the best solutions for implementing spacing in educational contexts would be the design of a *spiral curricula* in which contents are not taught just once, but instead, they are repeated over months and across courses (Kang, 2016). Thus, firstly, teachers should determine and stress in class what skills and knowledge are going to be needed in the future (Roediger & Karpicke, 2006a); and there should be an individualized SRS to show future teachers the materials which need to be revisited.<sup>13</sup> Testing techniques such as free recall, short answer questions, or fill-in-the-blank

---

<sup>13</sup> Some easier option would be to stick to the same teachers for a same class across courses. That way, teachers would have a better control on repetitions of a given material.

quizzes should be used on a regular basis, encouraging students to be constantly engaged in contents (Pashler, Bain, et al., 2007, Roediger & Karpicke, 2006a). Likewise, homework assignments should be used to practice previous learnt materials.

Some other methods that could be used are *interleaving* or *Peer Instruction* (PI). Interleaving is based on the repetition of materials from previous units in the review at the end of each unit or in pieces of homework. This strategy has had proven efficacy in educational contexts with standard curricula (Kang, 2016). Regular study books have exercises at the end of each unit related only to the given unit; interleaving promotes interleaving those exercises across different units in order to space them (Carpenter et al., 2012). According to Lasry (2008), PI is an instructional method created at Harvard by Eric Mazur. The development of any class depends on the real-time feedback given by students; typically, teachers explain a topic and then students have to solve a multiple-choice question after discussing it with their partner. Depending on the number of correct answers, the teacher will keep on explaining another topic or re-explain the previous one. Instant feedback can be given via a show of hands, flashcards, wired communication systems, wireless clickers, or smartphones.<sup>14</sup> This teaching method do not include free-recall or cued-recall tests, which had been shown to produce better results than multiple-choice questions (Kang et al., 2007), but it does keep the students engaged and challenged.

These methods and techniques will not be a universal remedy for classroom teaching, and maybe, only motivated students will benefit from them (Dunlosky et al., 2013). However, when used properly, they can enhance learning goals. For instance, even if students cannot recall some material in a delayed test, the mere fact of trying to retrieve it from memory will make them acquire this material in a much faster rate (Carpenter et al., 2012)

---

<sup>14</sup> PI was developed in the early 1990s so feedback mechanisms have been evolving according to advances in technology. Clickers are gadgets which include several buttons to press when a question is being asked and answers can be automatically displayed on the screen.

## 3.2. Implementation in the classroom

### a. Objectives

There is plenty of evidence supporting spacing and testing effects but not many studies have been conducted in the Spanish educational system. As stated in the introduction, these are the primary objectives of this experiment:

- To measure spaced retrieval effects in a real Spanish Compulsory Secondary Education classroom.
- To find the effects spaced retrieval practice can have in student's motivation.

### b. Context

The study was conducted in the I.E.S. Nicolás Salmerón y Alonso during an internship due to the study of a Masters in Secondary Teaching at the University of Almería. The subject area given by the intern was English language. Trainee's tutor, a native English speaker, was present in every instruction session given by the trainee. Intern's research and performance was approved at all times by the tutor.

### c. Participants

Thirty participants (13/14-years-old) were recruited from a second cycle at Compulsory Secondary Education (E.S.O.). All of them were participants in a *content and language integrated learning* (CLIL) class and their target language was English. Students were alphabetically divided in 2 groups (15 subjects vs. 15 subjects). Data from five participants were dropped as they could not assist to one or more study/testing sessions; the final sample size was of 25 participants, resulting in a minor difference between both groups (13 subjects vs. 12 subjects).

### d. Methodology

Students were taught for 5 different study sessions by the same trainee. In the first study session, a table with 103 food-related vocabulary words in English (Appendix

A) was distributed to every student. The vocabulary list was included in the curriculum and it was supposed to fit their level. Students were told to circle every unknown word and once they were done, the teacher explained every unfamiliar word and made the students write down their translations. Many of the words were unknown but some of them were not. Exposure to the study material lasted 30 minutes. Four days after that, three separate free recall tests with feedback were done for 53 words (16-19-18). Free recall means that students had to try to retrieve as many words as they could for each separate test, learning materials being left unseen. Nonetheless, students did not seem to pay much attention to feedback. The remaining 50 words were not tested.

A first group had the same test (*model A*) 6 and 26 days after the initial study, while the second group had a control test (*model B*) 6 days after the initial study and the same test (*model A*) as the first group 26 days after the initial study (A-A vs. B-A).

Two different tests were designed, each one with 20 different words: model A and model B (Appendix B). Model A had 11 words from the free recalled ones (4<sup>th</sup> day after the initial study) and 9 words from the non-tested word group. Model B had a mix of food-related words. In each model, 10 words had to be translated from English to Spanish and another 10 had to be translated from Spanish to English.

Before each test session, students were explained that tests would not count as a mark and that tests were part of a research so they should not try to copy or the results would not fit reality. Also, they were told that they had 10 minutes to do it. Then, tests were handed out. Every student finished the test before the time was over. Student's questions were not answered during the test. When everyone had finished, correct answers were handed out individually and they were given enough times to check their errors.

In order to check the degree of difficulty of the different translation directions, each word group had most of its words in a different direction than the other. Thus, in the 11-word group in model A, 8 words had to be translated from Spanish to English and 3 from English to Spanish (8 S→E; 3 E→S); while in the 9-word group in model A, only 2 words from Spanish to English and 7 words from English to Spanish (2 S→E; 7 E→S).

During corrections, being English the target language, English misspells (e.g. *curri*) were counted as errors whereas Spanish misspells (e.g. *berengena*) were not.

After the final test, satisfaction surveys (Appendix C) were provided with the aim of measuring the teacher’s intervention as part of a Masters’ internship. Questions were asked in Spanish for better understanding. From questions 1 to 8, students had to give a 1 to 5 mark; 1 meaning to strongly disagree and 5 to fully agree. Questions from 9 to 11 had an open answer. The survey sections related to the use of spaced retrieval were these: (1) *I found the classes enjoyable*; (7) *I feel we have made the most of our time in class*; (8) *overall, the teacher’s intervention has been satisfactory*; (9) *the thing I have liked the most of the teacher’s intervention has been...*; (10) *the thing I have liked the least of the teacher’s intervention has been...*; and (11) *further comments or suggestions*.

e. Results and discussion

Different learning distributions and techniques have been studied. The only test model assessed have been model A; model B served only as control test. Model A had two groups of words: the ones which had already been tested (11 words) via free recall and the ones which had not (9 words). In the end, there were a total of 4 different tested lags (Table 1).

Group	No. of students	No. of words tested	Days				Intervals		% Correct	
			1 <sup>st</sup>	5 <sup>th</sup>	7 <sup>th</sup>	27 <sup>th</sup>	Interstudy (days)	Retention (days)	7 <sup>th</sup> day	27 <sup>th</sup> day
1	13	11	Study session	Free recall	Cued recall	Final test	0-4-2	20	75.1	67.1
		9		-	Cued recall		0-6	20	73.5	70.0
2	12	11	Study session	Free recall	-	Final test	0-4	22	-	54.5
		9		-	-		0	26	-	59.3

Table 1. Schedules used in each of four different learning intervals and percentage correct on model A cued recalled test and final test. Source: original.



In general, the 11-word group appeared to be more difficult (54.5% correct answers in the final test) than the 9-word group (59.3% correct answers), although the 11-word group had been free recalled. Group 1's final retention had similar differences between the 11-word group and the 9-word group (2.9% vs. 4.8%). This could simply mean that the 11-word group words were more difficult, that Spanish-to-English translations are more difficult, or that free recall does not provide long-term benefits on its own.

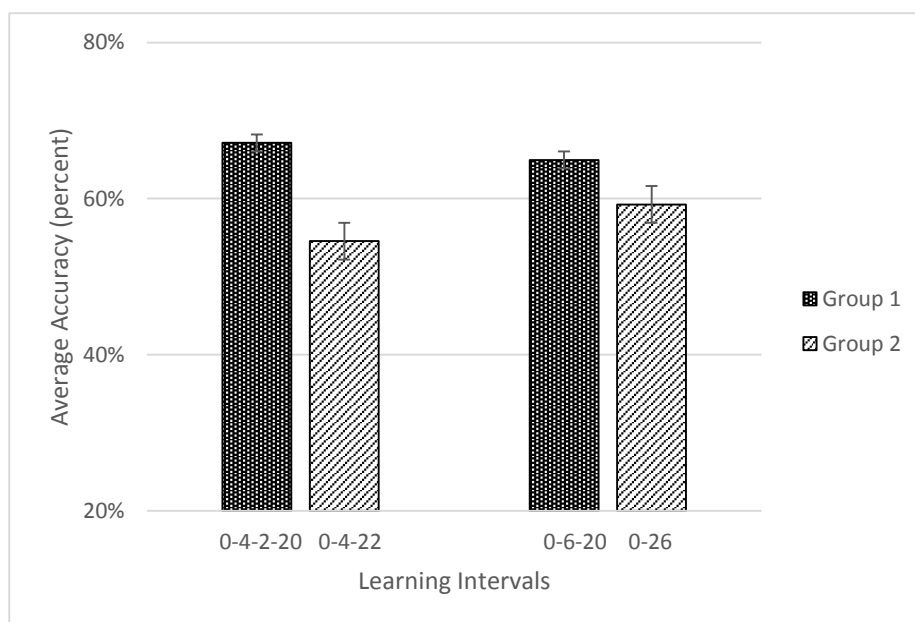


Figure 5. Final retention for every learning interval studied. On the left, the 11-word group studied by either group 1 or 2; on the right, the 9-word group studied also by both groups. Source: original.

Spaced retrieval (0-6-20) promoted higher retention in the final test (Figure 5) as compared to massed practice (0-26). Average accuracy was 10.7% higher for the spaced retrieval group. Although according to recent literature, higher long-term results were expected, in this research many variables could not be controlled and results still support spaced retrieval instead of massed learning. It is surprising the low difference between both groups given that the massed practice group had a 26-days RI and the spaced retrieval group had only a 20-days RI.

Particularly, for the 11-word group, adding a cued recall test 2 days after a first free recall one (0-4-2 ISI), seems to have improved retention in the final test by 12.6% as compared to just having a free recall test (0-4 ISI). The RI for the added cued recall interval is 2 days shorter than the just free recall interval and could explain the great difference.

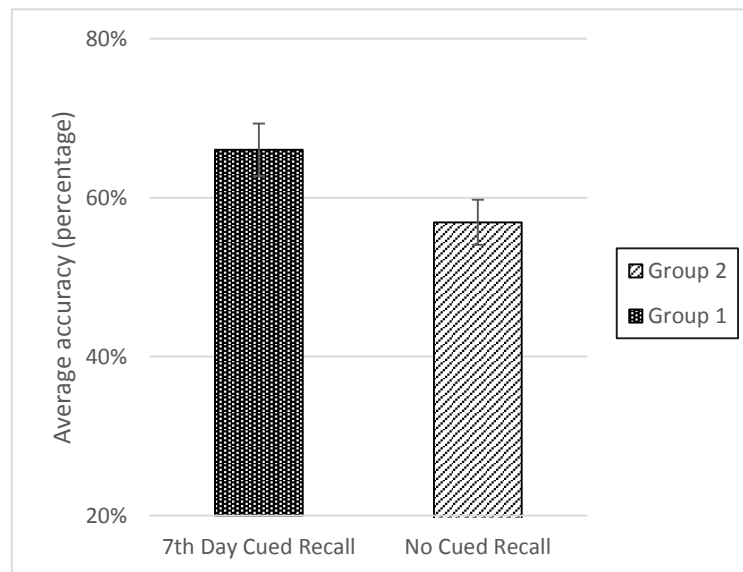


Figure 6. Cued recall effects 26 days after first study session. Source: original.

Unexpectedly, using just a cued recall study test instead of both a free and a cued recall study tests, appears to slow forgetting. The word group with the 0-4-2 ISI suffered a decreased in the accuracy level by -8% while the 0-6 ISI word group had only a -3.5% decrease for the same retention interval (20 days). This steeper decrease may be due to the proximity to a previous free recall test; thus, data could be interpreted as if free recall tests promoted short-term retention but not long-term retention. This could also fit the desirable difficulty theory (Karpicke & Roediger, 2007), which indicates that a longer ISI could promote higher long-term retention; the retrieval degree of difficulty was higher in the group with the 0-6 ISI and learners had to spend more time trying to retrieve previously studied knowledge.

Overall, cued recall testing showed to improve long-term retention – 26 days after the first study session – by 9.1% as compared to non-cued recall intervals (Figure 6). Cued recall tests were finished in about 5 minutes and 2 more minutes for feedback

were required; results are impressive given the short time this technique requires. This technique seems to slow forgetting,

With respect to satisfaction surveys, they show that testing practices on a regular basis do not seem to demotivate students. The option *I found the classes enjoyable* got an average of 4.66/5, close to fully agreement. Question number 7, *I feel we have made the most of our time in class*, also found great approval (4.58/5) and especially, *overall, the teacher's intervention has been satisfactory* (4.96/5). The rest of the sections showed 14 commentaries highlighting how entertaining lessons were, 5 positive remarks on how much have they learnt, and 2 comments exclusively dedicated to how much they have liked reviewing vocabulary. Most of the students left *the thing I have liked the least...* section either in blank or they specifically wrote there was nothing they had disliked. Results were similar for both genders; the maximum point's difference for a given question was 0.15 and proposals were also similar.

Although data could not be as controlled as in labs, results were similar as the ones found in recent literature. In this experiment, spaced retrieval practice outperformed massed practice, cued recall tests seemed to slow forgetting when compared to free recall tests, and using spaced retrieval practice on a daily basis did not appear to reduce in-class student's motivation.

#### f. Limitations of the experiment

First of all, the trainee did not receive formal specific instruction on how to properly conduct an experiment. In addition, this experiment was part of an internship so there were many limitations. There was only a short period of time to put the experiment into practice; five were the maximum number of lessons the trainee was allowed to give to the same class. As some other materials had to be taught, the time to conduct the experiment was limited; thus, the number of testing questions had to be reduced, results being less reliable than with a larger sample.

The followed methodology was not well planned. Uncontrolled free recall testing (recommended by the trainee's tutor) as a study method before test trials altered results in subsequent trials. Furthermore, free recall tests caused the formation of some

different lags than the ones originally planned; 1-4-2, 1-4, and 1-6 interstudy intervals were studied instead of larger 1, and 1-6 ISIs samples, which would have implied more reliable results. Changing interstudy intervals also altered retention intervals (20, 22, and 26 RIs as opposed to just 20 and 26 RIs). Additionally, the first study session did not show a Spanish translation of the material, students had to check if they knew the words and then ask for the translation of the ones they did not understand; this type of practice could have implied using the same cognitive mechanisms as in retrieval practice.

Finally, it is important to mention that lab experiments' study sessions are supposed to be longer for massed than for spaced practice. In this research experiment, massed and spaced initial study sessions were the same; thus, spaced practice inferred a longer study time of the same material compared to massed practice. This is due to the added difficulty of dividing a class in two groups when teaching real curriculum.

## 4. Conclusions

Spaced retrieval is a learning technique which comprises spacing and testing effects. Both of these techniques have been chosen as the learning techniques with the highest utility in recent reviews (Dunlosky et al., 2013; see also Pashler, Bain, et al., 2007). They have been studied for over a century (Ebbinghaus, 1885/1913; Gates, 1917) and they have shown to enhance learning not only of facts and lists, but also of complex theoretical concepts, motor skills, or in-class subjects (Smolen et al., 2016). In a large meta-analysis, Cepeda et al. (2006) found that, overall, spacing materials out instead of massing them, produced 15% higher retention, and it seemed to apply to people of all ages.

With respect to the best interval schedule during spacing to assure the highest retention, it looks like it depends on the time knowledge is going to be needed. According to Cepeda et al. (2008) and Pashler, Rohrer, et al. (2007), the best interstudy interval is about 10–20% of a given retention interval.<sup>15</sup> Thus, if information will be needed in 1 week, the ISI should be of 1 day, and if it will be needed in 1 year, the ISI

---

<sup>15</sup> See Figure 1 for an explanation of the terms interstudy interval and retention interval.

should be of about 2 months. Furthermore, there is a debate on whether spaced intervals should be expanding (e.g. 1-4-10) or equally spaced (e.g. 5-5-5). In general, reviews have not found a big difference between both approaches, similar long-term retention have been stated (Balota et al., 2007) or insufficient data (Carpenter et al., 2012; Cepeda et al., 2006).

On the other hand, memory retrieval is often used in conjunction with tests. Testing can bring mediated effects (e.g. higher study time when feedback is given) and direct effects (Roediger & Karpicke, 2006a). The testing effect usually refers to the direct ones; this means that the mere fact of recalling a studied material improves long-term retention as compared to spending more time exposed to the same material (Hopkins et al., 2015; Van Gog & Sweller, 2015). In Karpicke and Roediger (2007), testing showed 21% higher retention compared to restudying in a 1-week delay test after learning the study contents; these results are impressive given that feedback was not provided and also that restudy groups' exposure to the learning material was four-fold higher (Karpicke & Roediger, 2007).

Combining spaced learning and testing yields to even greater retention than using just one of them (Hopkins et al., 2015; Kang, 2016), this is why there have been several attempts to automatize spacing. The Leitner system is one of the first examples of this; it popularised flashcards spacing (see Figure 4). The SuperMemo algorithm (Wozniak, 1990) has served as a basis for recent software. Anki (Ankirs, 2016) is the most famous spaced repetition system (SRS) at the moment, but at the same time, in educational contexts, student's lack of motivation have been shown to interfere with its usage (Pérez Chamorro et al., 2015).

As mentioned before, spaced retrieval practice has proven its efficacy in educational settings, but still, this learning technique is not used by most of the teachers. The experiment presented in this essay have shown that spaced retrieval practice can be adapted to English teaching in secondary education using real curricula. Spaced retrieval practice lead to 11% higher retention 26 days after first study session compared to massed study.<sup>16</sup> Cued recall intervals outperformed non-cued recall ones by a 9%

---

<sup>16</sup> See Table 1.

increase in retention (see Figure 6). Also, using just a cued recall test instead of a free recall and a cued recall one in conjunction seems to slow forgetting. Finally, student's motivation remained high and, overall, the extended opinion was that lessons using spaced retrieval as a learning practice (and not as a measure tool) were highly enjoyable.

To conclude, even if the cognitive mechanisms behind spaced retrieval practice are not well understood, there is plenty of evidence showing its benefits. It is obvious that spaced retrieval practice has long been forgotten in our educational systems and its use on a daily basis is far from being a reality. This is why educational institutions should train teachers on the benefits of these techniques and on how they can be adapted to the classroom. What is more important, the implementation of a spiral curricula should be studied; thus, important contents would be repeated across courses over years (Kang, 2016), instead of dropping them from further practice soon after teaching them for the first time, as it usually occurs.

# References

- AMERICAN PSYCHOLOGICAL ASSOCIATION. (2010). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.
- ANKISRS. (2016). *Anki Manual*. [online] Available at: <http://ankisrs.net/docs/manual.html#spaced-repetition> [Accessed 8 Jun. 2016].
- ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed system and its control Processes. *Psychology of Learning and Motivation*, 2, 89-195.
- BACON, F. (2000). *Novum organum* (L. Jardine & M. Silverthorne, Trans.). Cambridge, England: Cambridge University Press. (Original work published 1620)
- BALOTA, D. A., DUCHEK, J. M., & LOGAN, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. *The foundations of remembering: Essays in honor of Henry L. Roediger, III*, (pp 83-105).
- BADDELEY, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423.
- BENIGAS, J. E. (2015). Spaced retrieval training: 26 years of growth. *SIG 15 Perspectives on Gerontology*, 20(1), 34-43.
- BUTLER, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-1133.
- CARPENTER, S. K. (2009). Cue Strength as a Moderator of the Testing Effect: The Benefits of Elaborative Retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563-1569.
- 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.

- CARPENTER, S. K., & OLSON, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 92–101.
- CARPENTER, S. K., PASHLER, H., WIXTED, J. T., & VUL, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448.
- 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.
- 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.
- DAVEY, E. (2014). *Spaced Learning applied to teaching biology*. (Unpublished master's thesis). Gävle University College.
- 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.
- EBBINGHAUS, H. (1913). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius Trans.) New York, NY: Teachers College, Columbia University. (Original work published 1885)
- ENDRES, T., & RENKL, A. (2015). Mechanisms behind the testing effect: an empirical investigation of retrieval practice in meaningful learning. *Frontiers in psychology*, 6.
- GATES, A. I. (1917). Recitation as a Factor in Memorising, No. 40. *Archives of Psychol., New York*.
- GOOGLE SCHOLAR. (2016). *Google Scholar*. [online] Available at: <https://scholar.google.com/> [Accessed 21 Jun. 2016].



- HOPKINS, R. F., LYLE, K. B., HIEB, J. L., & RALSTON, P. A. (2015). Spaced Retrieval Practice Increases College Students' Short-and Long-Term Retention of Mathematics Knowledge. *Educational Psychology Review*, 1-21.
- JAMES, W. (1890). *The principles of psychology*. New York: Holt.
- KANG, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, 3, 12-19.
- KANG, S. H., LINDSEY, R. V., MOZER, M. C., & PASHLER, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval?. *Psychonomic bulletin & review*, 21(6), 1544-1550.
- KANG, S. H., McDERMOTT, K. B., & ROEDIGER, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528-558.
- KARPICKE, J. D., & AUE, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, 27(2), 317-326.
- KARPICKE, J. D., & ROEDIGER, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704-719.
- KARPICKE, J. D., & ROEDIGER, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968.
- KUHL, D., & MAROLT, M. (2015). Mobile collaborative learning using flashcards on smartphones. *6th Postgraduate International Conference on Advancements in Computer Science and Applications intelligence*, 41-48.
- LANDAUER, T. K., & BJORK, R. A. (1978). Optimum rehearsal patterns and name learning. *Practical aspects of memory*, 1, 625-632.
- LASRY, N. (2008). Clickers or flashcards: Is there really a difference?. *The Physics Teacher*, 46(4), 242-244.
- LEITNER, S. (1984). *Así se aprende: psicología aplicada del aprender-camino al éxito* (pp. 68-80). Barcelona: Herder.

- MCDANIEL, M. A., FADLER, C. L., & PASHLER, H. (2013). Effects of spaced versus massed training in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1417-1432.
- MCDANIEL, M. A., THOMAS, R. C., AGARWAL, P. K., MCDERMOTT, K. B., & ROEDIGER, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, 27(3), 360-372.
- MNEMOSYNE-PROJ. (2016). *The Mnemosyne Project*. [online] Available at: <http://mnemosyne-proj.org/> [Accessed 8 Jun. 2016].
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533.
- MOSALINGUA. (2016). *The MOSA Learning Method*. [online] Available at: <http://www.mosalingua.com/en/our-apps/the-mosa-learning-method/> [Accessed 8 Jun. 2016].
- 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*.
- PASHLER, H., ROHRER, D., CEPEDA, N. J., & CARPENTER, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic bulletin & review*, 14(2), 187-193.
- PÉREZ CHAMORRO, V. A., GARCÍA ÁLVAREZ DE PEREA, J., & CASASOLA BALSELLS, M. A. (2015). Las tarjetas digitales (e-flashcards) en el aprendizaje autónomo de contabilidad financiera. *IJERI: International Journal of Educational Research and Innovation*, (4), 150-161.
- PIMSLEUR, P. (1967). A memory schedule. *The Modern Language Journal*, 51(2), 73-75.
- PIMSLEUR.COM. (2016). *Pimsleur Method*. [online] Available at: <http://www.pimsleur.com/pages/pimsleur-about-us> [Accessed 13 Jun. 2016].

- ROEDIGER, H. L., & KARPICKE, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181-210.
- ROEDIGER, H. L., & KARPICKE, J. D. (2006b). Test-enhanced learning taking memory tests improves long-term retention. *Psychological science*, 17(3), 249-255.
- ROWLAND, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463.
- SANTROCK, J. W. (2011). El enfoque del procesamiento de información. *Psicología de la Educación* (pp. 97-136). McGraw-Hill Interamericana de España.
- SMOLEN, P., ZHANG, Y., & BYRNE, J. H. (2016). The right time to learn: mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, 17(2), 77-88.
- SOBEL, H. S., CEPEDA, N. J., & KAPLER, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, 25(5), 763-767.
- SPITZER, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641.
- VAN GOG, T., & SWELLER, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, 27(2), 247-264.
- WIKIPEDIA. (2016). *Leitner system*. [online] Available at: [https://en.wikipedia.org/wiki/Leitner\\_system](https://en.wikipedia.org/wiki/Leitner_system) [Accessed 16 Jun. 2016].
- WING, E. A., MARSH, E. J., & CABEZA, R. (2013). Neural correlates of retrieval-based memory enhancement: An fMRI study of the testing effect. *Neuropsychologia*, 51(12), 2360-2370.
- WOZNIAK, P. A. (1990). Optimization of learning. (Unpublished master's thesis). Poznan University of Technology.
- WOZNIAK, P. A., & GORZELANCZYK, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta neurobiologiae experimentalis*, 54, 59-59.

# LIST OF FIGURES AND TABLES

Figure 1. The basic scheme of a spacing experiment.....	6
Figure 2. Recognition of trivia facts.....	9
Figure 3. Re-studying vs. retesting short passages. ....	13
Figure 4. The Leitner system. ....	17
Table 1. Schedules studied .....	23
Figure 5. Retention for every learning interval .....	24
Figure 6. Cued recall usage effects.....	25

## Appendix A: Study sessions material

Meat	Vegetables	Fruit	Fish	Dairy Products	Snacks	Breakfast Foods	Other Foods
beef	tomato/ tomatoes (Is it fruit or veg?)	apple	cod	milk	nuts	toast	sugar
pork	avocado	orange	haddock	full-fat milk	olives	jam	salt
chicken	peppers	banana	plaice	semi-skimmed milk	hummus	marmalade	pepper
bacon	mushroom	pear	tuna	skimmed milk	crisps	breakfast cereal	vinegar
ham	potato/ potatoes	peach	salmon	butter	chocolate	cornflakes	ketchup
lamb	peas	lemon	trout	margarine	sweets	muesli	mayonnaise
liver	carrot	lime	mackerel	cream	biscuits	porridge	mustard
kidneys	onion	plum	herring	cheese			spices
turkey	celery	melon	sardine	yoghurt			garlic
duck	cabbage	grape	anchovy				chilli
sausages	broccoli	mango					curry powder
cooked meat	cauliflower	kiwi fruit					pasta
salami	Brussels sprouts	apricot					pizza
pork pie	beans	pineapple					spaghetti
sausage rolls	leek	blackberry					eggs
paté	aubergine	blackcurrant					rice
	cucumber	strawberry					noodles
	courgette	raspberry					
	lettuce						

# Appendix B: Study tests (with answers)

## Model A

### Translate into Spanish:

1. Skimmed milk
2. Aubergine
3. Blackberry
4. Raspberry
5. Courgette
6. Cabbage
7. Avocado
8. Tuna
9. Pear
10. Pepper

### Translate into Spanish:

1. Skimmed milk – Leche desnatada
2. Aubergine – Berenjena
3. Blackberry – Mora
4. Raspberry – Frambuesa
5. Courgette – Calabacín
6. Cabbage – Col / Repollo
7. Avocado – Aguacate
8. Tuna – Atún
9. Pear – Pera
10. Pepper – Pimienta/o

### Translate into English:

1. Ajo
2. Mostaza
3. Curry en polvo
4. Albaricoque
5. Patatas fritas de bolsa
6. Bacalao
7. Cordero
8. Ciruela
9. Gachas de avena
10. Trucha

### Translate into English:

1. Ajo – Garlic
2. Mostaza – Mustard
3. Curry en polvo – Curry powder
4. Albaricoque – Apricot
5. Patatas fritas de bolsa – Crisps
6. Bacalao – Cod
7. Cordero – Lamb
8. Ciruela – Plum
9. Gachas de avena – Porridge
10. Trucha – Trout

## Model B

### Translate into Spanish:

1. Cook (person)
2. Course (restaurant)
3. Plate
4. Fancy (adj.)
5. Tip
6. Celery
7. Blackcurrant
8. Kidneys
9. Zucchini
10. Cream

### Translate into Spanish:

1. Cook (person) – Cocinero
2. Course (restaurant) – Plato
3. Plate – Plato
4. Fancy (adj.) – Sofisticado / Caro
5. Tip – Propina
6. Celery – Apio
7. Blackcurrant – Grosella negra
8. Kidneys - Riñones
9. Zucchini – Calabacín
10. Cream – Nata

### Translate into English:

1. Cuchillo
2. Camarero
3. Puerro
4. Entrante (2 options)
5. Entrante (2 options)
6. La cuenta
7. Caballa
8. Merluza
9. Cerdo
10. Leche entera

### Translate into English:

1. Cuchillo – Knife
2. Camarero – Waiter
3. Puerro – Leek
4. Entrante (2 options) - Appetizer
5. Entrante (2 options) - Starter
6. La cuenta – The bill
7. Caballa – Mackerel
8. Merluza – Haddock
9. Carne de cerdo – Pork
10. Leche entera – Full-fat milk

# Appendix C: Satisfaction survey

## Encuesta de satisfacción

Instituto: ..... Idioma: .....

Curso: ..... Sexo: M.... F.... Fecha: .....

Da tu opinión sobre los aspectos siguientes usando esta escala:

*5 = muy de acuerdo 4 = de acuerdo 3 = indiferente 2 = poco de acuerdo*

*1 = muy poco de acuerdo*

1. Las clases me resultaron entretenidas (...)
2. Las clases me resultaron fáciles (...)
3. Las explicaciones me resultaron claras (...)
4. Entendía al profesor perfectamente cuando hablaba en inglés (...)
5. La evaluación mediante presentación oral (teatro) me ha parecido justa (...)
6. Estoy satisfecho con el uso de los recursos electrónicos en el aula (...)
7. Siento que aprovechamos el tiempo en clase (...)
8. En general, la intervención del profesor ha sido satisfactoria (...)

9. Lo que MÁS me ha gustado del profesor en prácticas ha sido:

.....  
.....  
.....  
.....

10. Lo que MENOS me ha gustado del profesor en prácticas:

.....  
.....  
.....

11. Otros comentarios o sugerencias:

.....  
.....  
.....