REVISTA SIGNOS. ESTUDIOS DE LINGÜÍSTICA ISSN 0718-0934
© 2019 PUCV, Chile 52(100) 290-305

Immediate testing is more beneficial than delayed testing when learning novel words in a foreign language*

Las pruebas inmediatas son más beneficiosas que las pruebas diferidas cuando se aprenden palabras nuevas en una lengua extranjera

Roberto Andrés Ferreira
UNIVERSIDAD CATÓLICA DE LA SANTÍSIMA CONCEPCIÓN
CHILE
rferreira@ucsc.cl

Valeska Soto Sierra
UNIVERSIDAD CATÓLICA DE LA SANTÍSIMA CONCEPCIÓN
CHILE
vgsoto@emingles.ucsc.cl

Stephanie Aedo Vega
UNIVERSITY OF ESSEX
REINO UNIDO
sfaedo@emingles.ucsc.cl

Recibido: 09-XII-2017 / **Aceptado:** 24-IX-2018
DOI: 10.4067/S0718-09342019000200290

Abstract

The benefit of testing on the retention of verbal materials has been studied quite extensively, however very little attention has been put on establishing when the best time to actually apply a test is. The present study investigated the effect of testing (immediate and delayed) on the learning of novel words in English as a foreign language (EFL). The participants were 20 students of EFL enrolled on a 5-year teaching programme. They learned the meaning of 20 matched novel words presented with images, sentences, and exercises during learning. The experiment took place over a week. On day 1 participants learned a list of 10 words, and a day later (day 2), they learned another set of 10 words and were then immediately tested on all 20 words. On day 8, participants were tested again on all the words they had learned. A semantic categorisation task was used for the purpose, consisting of classifying newly learned words into living or nonliving things. The results showed that participants classified more accurately and responded faster to newly learned words tested immediately after training than words learned a day earlier, and these effects were stable over time. These results can be explained by interference theories or by the alternative retrieval route theory because when testing is applied immediately after learning, it acts as an instant shield that protects newly learned words from interference, or strengthens their retrieval routes.

Key Words: Testing effect, word learning, foreign language learning, proactive interference, semantic categorisation.

Resumen

El beneficio de las pruebas en la retención de estímulos verbales se ha estudiado ampliamente, sin embargo, se ha puesto muy poca atención en establecer el mejor momento para aplicar una prueba. El presente estudio investigó el efecto de las pruebas (inmediatas y diferidas) en el aprendizaje léxico en inglés como lengua extranjera (ELE). Los participantes eran 20 estudiantes de ELE de un programa de pedagogía de 5 años. Ellos aprendieron el significado de 20 palabras nuevas presentadas con imágenes, oraciones y ejercicios. El experimento tuvo lugar durante una semana. El día 1 los participantes aprendieron una lista de 10 palabras, un día más tarde (día 2) aprendieron otro conjunto de 10 palabras, y luego se les evaluó su conocimiento sobre las 20 palabras. El día 8, los participantes fueron evaluados nuevamente en todas las palabras que habían aprendido. Se utilizó una tarea de categorización semántica para este propósito, que consistía en clasificar las palabras recién aprendidas en seres vivos u objetos. Los resultados mostraron que los participantes clasificaron más correctamente y con mayor rapidez las palabras que fueron evaluadas inmediatamente después del entrenamiento que las palabras aprendidas el día anterior, y estos efectos se mantuvieron estables a lo largo del tiempo. Estos resultados pueden explicarse por teorías de interferencia o por la teoría alternativa de ruta de recuperación, ya que cuando las pruebas se aplican inmediatamente después del aprendizaje, actúan como un escudo instantáneo que protege las palabras recién aprendidas de la interferencia, o fortalece sus rutas de recuperación.

Palabras Clave: Efecto de las pruebas, aprendizaje de palabras, aprendizaje de lenguas extranjeras, interferencia proactiva, categorización semántica.

INTRODUCTION

Over the past few decades, the effect of testing on memory and learning has been studied quite extensively. Contrary to popular belief, testing is not simply a neutral event that should be used only to measure learning, but can in fact modify and enhance memory retention, as documented by a number of studies (Wheeler, Ewers & Buonanno, 2003; Roediger & Karpicke, 2006; Toppino & Cohen, 2009). In broad terms, taking a test on information learned during a study period has a positive effect on long-term retention in comparison with continued and repeated study (Roediger & Karpicke, 2006). It is clear that the consolidation function on memory of a process of testing contradicts the traditional approach to successful learning, which suggests that tests are simply a checkpoint for several study phases. In fact, testing can be responsible for the modification of memory and the enhancement of accessibility to target information.

The testing effect has been studied using mostly verbal materials (words) with emphasis on retrieval processes (Pyc & Rawson, 2010; Karpicke & Grimaldi, 2012; van den Broek, Takashima, Segers, Fernández & Verhoeven, 2013; Mulligan & Picklesimer, 2016; Cho, Neely, Crocco & Vitrano, 2017). There is evidence to suggest that testing protects memories since it prevents the build-up of proactive interference (Szpunar, McDermott & Roediger III, 2008; Weinstein, McDermott & Szpunar,

2011). Considering that the acquisition of verbal materials seems prone to interference, and that testing during encoding protects the consolidation of newly learned words, tests applied immediately after training should protect recent memories of words more effectively than if applied following a delay.

Most previous studies regarding testing have used single words in a first language (L1) as stimuli, with designs that have very little to do with how humans actually learn new vocabulary, including the memorization of word lists or word pairs. In the present study, we investigated the effect of testing time on the learning of novel words in English as a foreign language (EFL) using a more ecological design than in previous studies. More specifically, we wanted to know whether applying a test immediately after learning new words (accompanied by images and sentences) enhances retention over time to a greater extent than a delayed test a day after a study period.

1. Background

The first study on the effects of testing was that of Tulving (1967), which addressed the potential active role of testing in learning lists of words. At the time, little was known about the effects of prior recall tests on subsequent final retrieval, so he designed two experiments to compare the effects of prior presentations and prior recalls on subsequent recall of common nouns. The results showed that the number of words recalled depended on the total amount of time spent on the task, rather than the distribution of the time between studying and retesting the words, indicating that ultimately a test trial was as good as a study trial for improving learning of verbal material.

In more recent research, Tulving's findings have been extended. For instance, the effect of testing has commonly been observed in studies where time of exposure to the stimuli has been equally distributed between restudy conditions and testing conditions (Kornell, Hays & Bjork, 2009; Kang, 2010). These results have led to the conclusion that the beneficial effects of testing on retrieval do not depend on the time spent learning, but are directly tied to the act of testing itself.

There are some theories that have attempted to explain why tests are more beneficial than studying. For instance, Karpicke and Roediger (2007) argued that testing produces benefits to retention due to the engagement of retrieval processes while accessing information (e.g., words) stored in memory. Thus, the more an individual practices retrieval skills on initial tests, the better they will do on later retrieval instances, which in turn is reflected in enhanced performance in future tests (Roediger & Karpicke, 2006; Roediger & Butler, 2011). Unlike tests, repeated study only provides additional exposure to items without the need for explicit retrieval processes, which produces rapid initial learning, but poor long-term retention. By contrast, testing produces slower and more effortful initial learning, but results in better long-term retention and performance (Roediger & Karpicke, 2006; Halamish &

Bjork, 2011). Along these lines, Wheeler et al. (2003) investigated learning and forgetting rates under a repeated study condition –participants studied a list of words five times with no test trials– and a repeated study condition –participants studied the list of words only once and had four different recall tests trials. The rates of forgetting between these two types of encoding conditions revealed that after a short retention interval of 5 minutes, repeated study resulted in a higher level of recall of the target list; however, the lower rate of immediate recall obtained with the repeated test condition developed into a much lower rate of forgetting after a 7-day retention interval. Similarly, Karpicke and Roediger (2007) investigated long-term gains in learning a list of new words across several study phases and test trials, with a final recall test a week after learning. Participants were assigned to three different conditions: standard condition (study was alternated with test trials), repeated study (list of words were studied three times and there was only one recall test) and repeated testing (the list of words was studied only once and there were three consecutive recall tests). The results from this experiment showed that although learning curves for the three conditions were initially very similar, one week later the repeated testing group recalled significantly more words, indicating enhanced long-term retention.

The emphasis on retrieval processes is key to understanding memory and learning, as well as the long-term retention gains offered by testing. Based on this, Karpicke and Grimaldi (2012) argue for a retrieval-based perspective of learning that views retrieval as a reflection of the contents of knowledge constructed from previous encoding experiences. In this view, retrieval influences learning directly, given that every time knowledge is retrieved, it is altered, along with the ability to reconstruct it in the future, producing gains especially in long-term retention. Evidently, Karpicke and Grimaldi's (2012) perspective opposes traditional understandings of learning with emphasis only on the construction and storage of knowledge (Ausubel, 2012). The findings of the studies presented above are consistent in suggesting that tests not only assess learning but also greatly enhance it. When compared to repeated study conditions, the gains obtained by testing are typically not observed immediately after learning, but rather following a delay, which accounts for the long-term retention enhancement that testing provides (Toppino & Cohen, 2009).

All in all, the studies reviewed earlier converge in suggesting that testing is beneficial beyond repeated study. Testing involves information retrieval, so every time a test is applied, the future accessibility of that information improves because retrieving enhances the effectiveness of the specific cues involved in reconstructing all associated memories (Karpicke & Smith, 2012). More recently, neuroimaging evidence has also shown support for the testing effect (van den Broek et al., 2013; van den Broek, Segers, Takashima & Verhoeven, 2014). These studies suggest that testing generates changes in the connections within semantic networks in the brain, which enable the formation of additional associations and alternative routes of semantic

networks that selectively strengthen target responses, while inhibiting related but irrelevant ones.

As shown above, testing effects are well-established and studied using different methodologies. However, some aspects of testing are still not well-understood. For instance, there is no clear idea as to when testing is more efficient, and whether the effects seen in the study of verbal material in a first language can also be extended to foreign languages. Evidence from memory research suggests that retrieval might exert some sort of protection, as immediate retrieval of studied information blocks interference from other future information (Brown, Neath & Chater, 2007; Lewandowsky, Oberauer & Brown, 2009). Hence, if a test is applied immediately after learning, there are fewer possibilities for interference than if the test is applied a few hours later, in which case memory traces have been reactivated or regenerated through retrieval (Brown et al., 2007).

The present study investigated whether testing taking place immediately after learning new words results in improved retention over time compared with delayed testing (a day later). To the best of our knowledge, there are no studies that have addressed this issue during the learning of novel words in either L1 or L2. The only study that has shown evidence regarding the effect of tests immediately applied versus those applied after a delay has found that testing seems to be particularly beneficial when it comes to time responses. However, this benefit was found in both immediate and delayed test conditions when compared with repeated study (van den Broek et al., 2013). In the present study we hypothesized that testing immediately after learning would be better than after a delay, due to the protection and gains in memory strength that test retrieval provides (Brown et al., 2007; Halamish & Bjork, 2011). The sooner testing takes place, the less interference there is going to be on the consolidation of target novel words, resulting in better retention of their meanings, both in the short term and after a relatively long period of time.

2. Method

2.1. Participants

The participants were 20 students from the Universidad Católica de la Santísima Concepción (UCSC), Chile (8 males, 12 females; mean age 22.6 years; range 21-28) with normal hearing, normal or corrected-to-normal vision, and no language disorders or learning disabilities. They were all native speakers of Chilean Spanish, with at least three years of full-time formal instruction in English. The students who participated in the study had all passed a local examination that mimicked the FCE test (UCLES, 2016), and had all been classified as B2 level or above according to the Common European Framework of Reference for Languages (Council of Europe, 2016). Participants gave informed consent and received payment in exchange for their participation.

2.2. Stimuli

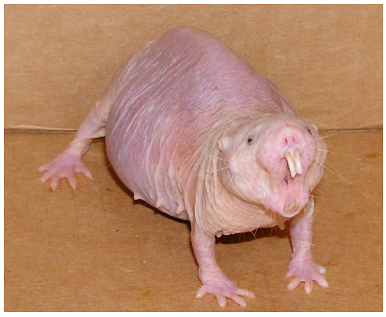
Two sets of 10 words that followed English phonotactic constraints were used in the experiment. The nonwords were matched on letter length, bigram frequency and mean RTs (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman, 2007) across experimental conditions (see Table 1).

Table 1. Log mean bigram frequency and mean RT for nonwords used in the experiment.

Experimental conditions		Semantic features	Letter length	Log mean bigram frequency	Mean RT
Immediate testing	Mean	14.3	5.4	3.2	789
	SD	4.9	1.2	0.2	63.8
Delayed testing	Mean	14.3	5.5	3.2	782
	SD	4.9	0.5	0.2	82.3

We paired the nonwords with real but obscure concepts corresponding to rare animals or rare objects, and obtained 10 living and 10 nonliving entities (e.g., *Sernal* was paired with the Phaistos Disc, which is a stamped disk of fired clay from the Minoan palace of Phaistos on the Greek of Crete, and *Chapice* was paired with the naked mole-rat, which is a rare mammal native to parts to East Africa) (see Figure 1 A and B). All novel words were concrete nouns and were accompanied by an auditory version recorded by a native male speaker.

A



B

Figure 1. Sample images used in the learning session. A. naked mole-rat. B. the Phaistos Disc.

We also created five sentences that described the meaning of each word, containing 14.3 features on average in each condition (see Table 2). The sentences were matched across conditions on the number of semantic features or attributes (e.g., is small, is made of clay, etc.) they provided (see Table 1). Semantic features have played an important role in constructing theories and models of semantic memory,

and have been widely used to study semantics (McRae, Cree, Seidenberg & McNorgan, 2005).

Table 2. Sample novel words and sentences used in the learning session.

Novel word	Sentences
chapice	The chapice lives in East Africa and is well adapted to its underground existence. A chapice has little hair and wrinkled pink or yellow skin. A chapice is small and has large protruding teeth. It digs and feeds on tubers. A chapice has two small eyes and its visual acuity is poor. A chapice has very short thin legs. It can move forward and backwards equally fast.
sernal	The sernal was discovered in Crete by an Italian archaeologist. The sernal dates back to the second millennium (B.C.E) and was probably used as a syllabary or an alphabet. The sernal is a unique item of no more than 15 cm in diameter. The sernal is made of clay and is covered with stamped symbols. The sernal's symbols represent people, tools, plants, and animals.

2.3. Procedure

The experiment used a repeated-measures or within-subjects design, so each set of words represented a different experimental condition (immediate or delayed testing), with the same participants taking part in the learning and testing of both sets of words. In this type of design, there is no need for a control group because each participant is its own control.

The experiment was conducted over the course of three days, including two consecutive days and an extra session eight days after initial exposure (see Figure 2).

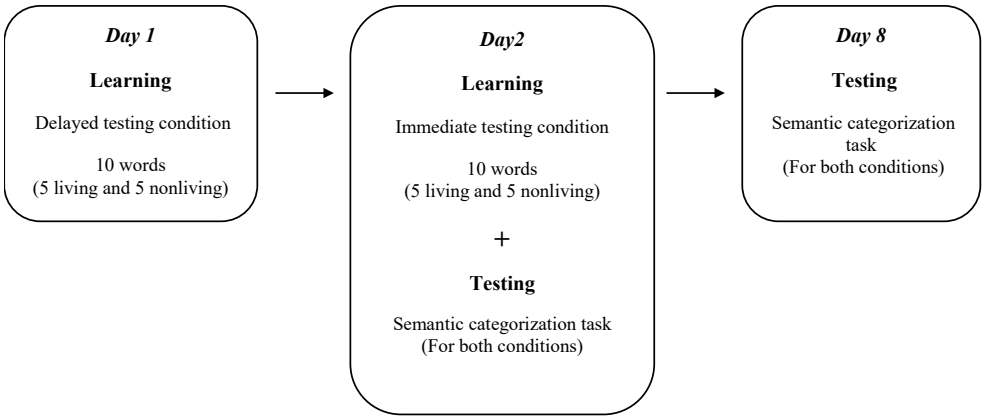


Figure 2. Schematic representation of the procedure.

2.3.1. Learning

Upon arrival, participants received a brief explanation of the purpose of the experiment and signed a written consent form. Each participant was exposed to the stimuli in a quiet room equipped with computers and headphones.

E-Prime software (Schneider, Eschman & Zuccolotto, 2002) was used to present all of the stimuli used across the learning sessions. The stimuli were distributed across two sets containing 10 novel words each, and each set was randomly assigned to either immediate or delayed testing conditions. On the first day, participants were presented with a set of 10 novel words and were asked to learn their phonological and orthographic forms, as well as their meaning. Each novel word was first presented on the screen (written form) accompanied by an image for 4,000 ms, and participants were required to say the word aloud after they heard it through the headphones. Each word was presented three times in this modality. Then all the novel words were presented embedded in sentences and accompanied by the same image displayed earlier. Each sentence presented a number of semantic features that describe the meaning of the target word. Participants were instructed to read each sentence and then press the space bar in order to move to the next sentence. At the end of each set of sentences, participants were asked a question about the information they had read in order to make sure they had been paying attention. Finally, they were presented with each novel word and an image, and had to re-type each word in order to learn their orthographic form (see Figure 3). The following day, participants were required to come back to the lab and go through the second and last learning session. They were exposed to 10 new words (5 living and 5 nonliving) and went through exactly the same procedure as the day before.

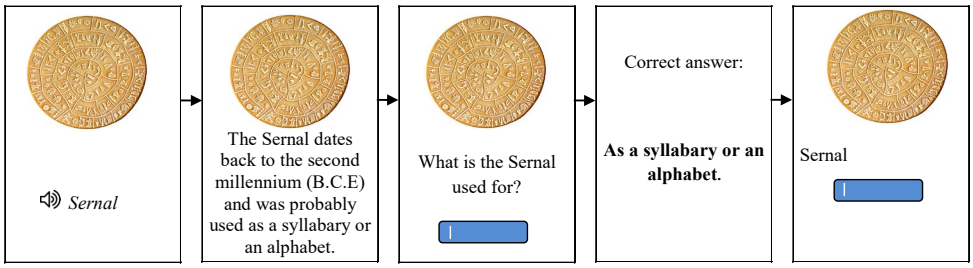


Figure 3. Schematic representation of the learning sessions.

2.3.2. Testing

After the second learning session, participants were tested on the novel words they had learned the day before (delayed testing) and on the same day (immediate testing). A semantic categorization task, in which participants had to classify each newly learned word as living or nonliving, was used. The task began with four practice trials

to enable participants to familiarise themselves with the procedure. Then the actual task began with the presentation of a fixation cross for 1,000 ms followed by a blank screen for 500 ms and a newly learned word (either living or nonliving) for 5,000 ms or until participants made a response, by pressing '1' for living and '2' for nonliving on the keyboard. They were asked to press the keys as fast and as accurately as they could. Accuracy rates and reaction times (RTs) were recorded. Finally, a week after initial exposure (on day 8), participants were asked to perform the same test again without being further exposed to the stimuli.

3. Results

The 20 participants in our study contributed 800 responses, of which 28.5% corresponded to miscategorisation errors. A summary of the results is presented in Table 3. We used mixed-effects models for accuracy and RT analyses because they provide the possibility of including:

> "subjects and items as crossed, independent, random effects, as opposed to hierarchical or multilevel models in which random effects are assumed to be nested" (Baayen, Davidson & Bates, 2008: 391).

Hence, linear mixed-effects models are more appropriate for analysing linguistic data with several observations by participants than other tests such as analysis of variance (ANOVA). All the analyses were conducted in R version 3.2.5. (R Core Team, 2016) using the lme4 package version 1.1-12 (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai, Grothendieck & Green, 2016). In order to report significance, we used the lmerTest package version 2.0-32 (Kuznetsova, Brockhoff & Christensen, 2016).

Table 3. Mean RTs in ms (with SDs) and error rates in the semantic categorization task as a function of testing time and day.

	Day 2	Day 8
	Immediate testing	
Mean RT	1230	1077
SD	383	360
% errors	20	23
	Delayed testing	
Mean RT	1372	1136
SD	442	407
% errors	36	35

3.1. Reaction Times (RTs)

A linear mixed-effects model was run on the RT data in order to test the effect of testing (immediate, delayed), day (Day 2, Day 8) and the interaction between the two factors. There was a main effect of testing ($p < .05$) and day ($p < .001$), but no interaction ($p = .25$). More specifically, participants were able to classify the words

faster when they were tested immediately after the learning session (immediate testing) than a day later (delayed testing). Both conditions showed better performance on Day 8 than on Day 2, but the advantage for the immediate testing condition was stable over time (see Figure 4 and Table 4).

3.2. Accuracy

We conducted a logistic linear mixed-effects model on the accuracy data. There was a significant effect of testing ($p < .01$), with newly learned words tested immediately after learning outperforming words tested on the following day. We did not find an effect of day ($p > .05$) or an interaction between testing and day ($p > .05$), which means that eight days after initial exposure, without the mediation of further training, the advantage for immediate testing over the delayed testing condition remained stable (see Figure 4 and Table 4).

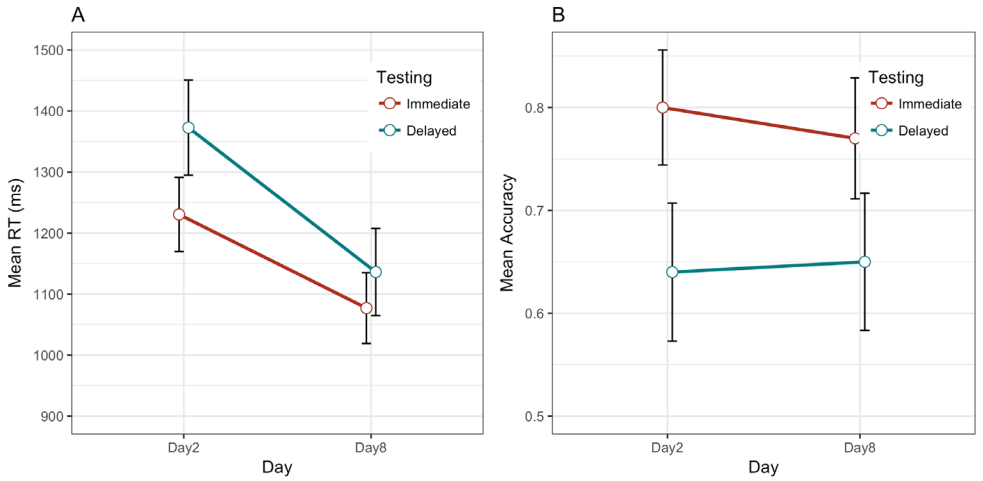


Figure 4. Results for RTs (A) and accuracy (B) in the semantic categorization task.

Table 4. Analysis of RT and accuracy. Coefficients of the main effects and interaction effects of the model together with the standard error (SE), t-values, and p-values in test and day.

	Estimate	**SE**	***t*-value**	***p*-value**
	Reaction Time (RT)			
Intercept	25.42	58.16	0.437	0.665
Testing_Delayed	141.81	50.62	2.801	0.008
Day_Day 8	-161.85	38.43	-4.211	0.000
Testing:Day	-65.99	57.24	-1.153	0.249
	Accuracy			
Intercept	0.80	0.04	17.97	0.000
Testing_Delayed	-0.16	0.05	-3.18	0.003
Day_Day 8	-0.03	0.04	-0.70	0.483
Testing:Day	0.04	0.06	0.66	0.508

4. Discussion

The aim of the present study was to assess the effect of immediate and delayed testing on the semantic categorisation of newly learned words in English as a foreign language (EFL). To the best of our knowledge, this is the first study that has compared immediate and delayed testing on the learning of novel words in a foreign language, and using a fairly ecological paradigm. We found a significant effect of testing, with words tested immediately after training showing an advantage over words that were tested a day later. Crucially, there was no interaction between testing and day, which means that the initial advantage for immediate testing was the same across day 2 and day 8. The effect of day was only significant for the RT data, which implies that response times were faster on day 8 for both testing conditions (immediate and delayed). Overall, these results mean that testing immediately after training is more beneficial for learning the meaning of new words in English as a foreign language than delayed testing. In the following paragraphs, we offer an explanation for these results based on interference and alternative retrieval route theories.

A number of researchers have proposed that testing can reinforce learning and protect new memories from the interference of other sources of information (Szpunar et al., 2008; Weinstein et al., 2011; Nunes & Weinstein, 2012). Therefore, the time at which participants are tested seems to be a good predictor of retention of explicit information about a word's meaning. Testing, in general, makes recall processes more efficient by allowing relevant information to be activated, and irrelevant information to be attenuated (Karpicke & Smith, 2012). Hence, it seems that testing promotes the development of refined mnemonic associations (Pyc & Rawson, 2010) and improves the accessibility to the encoding context, reducing forgetting over time (Wheeler et al., 2003; Carpenter, Pashler, Wixted & Vul, 2008). Furthermore, testing enriches semantic networks because additional associations and alternative retrieval routes are formed, and also refines memory representations because it selectively strengthens target responses while inhibiting related but irrelevant ones (Carpenter & DeLosh, 2006).

In the present study, we have demonstrated that all the above features of testing are enhanced if the test is applied immediately after learning new words in a foreign language, in comparison with a delayed application. More specifically, the immediate application of a test makes the protective effect of testing against the build-up of proactive interference more effective because it can instantly block new information from interfering with newly formed memories (Szpunar et al., 2008; Weinstein et al., 2011; Nunes & Weinstein, 2012). In words of Szpunar et al. (2008), testing is, in fact, useful not only for learning new material, but also for protecting newly acquired information from the intrusion of future content. This idea has also been supported by more recent research; for instance, Wahlheim (2015) argues that testing applied to initially learned information, before presenting new material, can reduce the effect of

proactive interference because it isolates competing sources of information. He further explains that testing promotes the integration of initially studied information, making it more accessible during the presentation of new content.

In our study, participants learned a list of words, waited for a day to learn a second list, and were finally tested on both lists of words. Keeping in mind that testing acts as a shield that protects memories from proactive interference, we propose that immediate testing here produced instant protection of the newly learned words from proactive interference coming from new information. In contrast, delayed testing could not protect newly learned words from interference during the day that passed between learning and testing, which resulted in poorer performance reflected in lower accuracy and slower response times. The fact that these results were found across day 2 and day 8 shows that immediate testing also affects the consolidation of newly learned words over time.

The current results are also supported by a more general interference theory, which states that forgetting occurs because recently formed memories have not yet had the chance to consolidate, so they are vulnerable to the interfering force of mental activity and mental information (Baddeley & Hitch, 1993). Given that testing requires a higher amount of mental effort, it provides a strengthening of semantic networks and thus interference is avoided and results in recently learned novel words remaining stable and consolidated in the long-term (Pyc & Rawson, 2009, 2010). In this view, immediate testing acts as a shield for new incoming information, and it is more effective than delayed testing, since it prevents newly formed memories from becoming less stable due to lack of instance reinforcement.

The facilitation of memory retrieval promoted by testing soon after exposure and leading to better performance in our study can also be attributed to the strengthening of alternative retrieval routes in semantic networks (Lockhart, 2002). If we follow this framework, the proposal goes that the sooner the retrieval routes are created, the easier it is to retrieve newly stored lexical items. Thus, in our delayed testing condition, we suggest that retrieval routes were established after a 24-hour period, which made them less successful in providing access to the meaning of the newly learned words.

It is true that the difference between immediate and delayed testing conditions on day 2 alone could be simply attributed to a general effect of recency, because if a test is applied immediately after learning, participants are more likely to remember that information than the information presented a day earlier. The effect of recency has been well-documented in the history of memory research; for instance, when participants are presented with a list of words, they are more likely to recall the last words of the list than those presented earlier. For Baddeley and Hitch (1993: 146), the recency effect "reflects the application of an explicit retrieval strategy to the residue of implicit learning within a range of cognitive systems", so it is this retrieval strategy that

allows participants to remember recent information better. In the current study, we suggest that immediate testing is not simply a general recency effect, because the difference between the conditions was stable over time (after a week), meaning that immediate testing was effective not only because participants could use strategies to classify words into living or nonliving things, but also because it could block interference from future information or strengthen retrieval routes, making memories more stable over time.

CONCLUSION

The present study has provided evidence that immediate testing is more beneficial than delayed testing for learning the meaning of novel words in a foreign language, and that its effects remain constant a week later. Since testing acts as a shield that protects newly learned words from interference and/or allows the creation of alternative retrieval routes, the sooner a test is applied the better the retention of newly learned words. In sum, immediate testing seems ideal to effectively consolidate newly formed memories over time and, in this particular case, to promote vocabulary learning. In practice, testing should be applied not as a punitive task, but as a way to protect the memories that support newly learned words from interference from future content that the students are exposed to during the day. Allowing students to retrieve the content they have just learned generates more stable memories over time, hence boosting word learning.

REFERENCES

Ausubel, D. P. (2012). *The acquisition and retention of knowledge: A cognitive view*. New York: Springer Science & Business Media.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baddeley, A. D. & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory* & *Cognition*, *21*(2), 146-155.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459 [on line]. Retrieved from: https://doi.org/10.3758/BF03193014

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G. & Green, P. (2016). *lme4: Linear mixed-effects models using "Eigen" and S4* (Version 1.1-12) [on line]. Retrieved from: https://cran.r-project.org/web/packages/lme4/index.html

Brown, G. D., Neath, I. & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539.

Carpenter, S. K. & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory* & *Cognition*, *34*(2), 268-276.

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 [on line]. Retrieved from: https://doi.org/10.3758/MC.36.2.438

Cho, K. W., Neely, J. H., Crocco, S. & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *The Quarterly Journal of Experimental Psychology*, *70*(7), 1211-1235.

Council of Europe. (2016). *Common European Framework of Reference for Languages* [on line]. Retrieved from: https://www.coe.int/en/web/common-european-framework-reference-languages/home

Halamish, V. & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801.

Kang, S. H. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory* & *Cognition*, *38*(8), 1009-1017.

Karpicke, J. D. & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*(3), 401-418.

Karpicke, J. D. & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151-162.

Karpicke, J. D. & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17-29.

Kornell, N., Hays, M. J. & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models (Version 2.0-32) [on line]. Retrieved from: https://cran.r-project.org/web/packages/lmerTest/index.html

Lewandowsky, S., Oberauer, K. & Brown, G. D. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*(3), 120-126.

Lockhart, R. S. (2002). Levels of processing, transfer-appropriate processing, and the concept of robust encoding. *Memory*, *10*(5-6), 397-403.

McRae, K., Cree, G. S., Seidenberg, M. S. & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547-559.

Mulligan, N. W. & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 938.

Nunes, L. D. & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory*, *20*(2), 138-154.

Pyc, M. A. & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437-447.

Pyc, M. A. & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335-335.

R Core Team. (2016). *A language and environment for statistical computting. R Foundation for Statistical Computing. Vienna, Austria* [on line]. Retrieved from: https://www.r-project.org/

Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255 [on line]. Retrieved from: https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Szpunar, K. K., McDermott, K. B. & Roediger III, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392.

Toppino, T. C. & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252-257.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*(2), 175-184 [on line]. Retrieved from: https://doi.org/10.1016/S0022-5371(67)80092-6

UCLES. (2016). *Cambridge English: First (FCE) | Cambridge English* [on line]. Retrieved from: http://www.cambridgeenglish.org/exams-and-tests/first/

van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G. & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, 78, 94-102 [on line]. Retrieved from: https://doi.org/10.1016/j.neuroimage.2013.03.071

van den Broek, G. S., Segers, E., Takashima, A. & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, *22*(7), 803-812.

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory* & *Cognition*, *43*(1), 27-38 [on line]. Retrieved from: https://doi.org/10.3758/s13421-014-0455-5

Weinstein, Y., McDermott, K. B. & Szpunar, K. K. (2011). Testing protects against proactive interference in face–name learning. *Psychonomic Bulletin* & *Review*, *18*(3), 518.

Wheeler, M., Ewers, M. & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571-580 [on line]. Retrieved from: https://doi.org/10.1080/09658210244000414

* ACKNOWLEDGEMENTS

This work was supported by CONICYT-Chile under Grant FONDECYT 11130678.