UNIVERSIDADE FEDERAL DE MINAS GERAIS

FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS

PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

TATIANA SALAZAR SILVA PINTO

THE RETRIEVAL PRACTICE: A THEORETICAL REVIEW AND AN EMPIRICAL STUDY

Belo Horizonte - MG

2018

TATIANA SALAZAR SILVA PINTO

THE RETRIEVAL PRACTICE: A THEORETICAL REVIEW AND AN EMPIRICAL STUDY

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais para obtenção do título de Mestre em Psicologia.

Área de concentração: Desenvolvimento Humano

Orientador**:** Prof**.** Dr. Antônio Jaeger

Belo Horizonte - MG

2018

150
P659r
2018

Pinto, Tatiana Salazar Silva

The retrieval practice [manuscrito] : a theoretical review and an empirical study / Tatiana Salazar Silva Pinto. - 2018.

56 f.

Orientador: Antônio Jaeger.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas.

Inclui bibliografia

1. Psicologia - Teses . 2. Psicologia aplicada - Teses. 3. Testes – Teses. 4. Aprendizagem – Teses. 5. Educação – Teses. I. Jaeger, Antônio. II. Universidade Federal de Minas Gerais. Faculdade de Filosofia e Ciências Humanas. III. Título.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

FOLHA DE APROVAÇÃO

Prática da evocação e habilidades de leitura em crianças do 6° ano.

TATIANA SALAZAR SILVA PINTO

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em PSICOLOGIA, como requisito para obtenção do grau de Mestre em PSICOLOGIA, área de concentração DESENVOLVIMENTO HUMANO, linha de pesquisa Cognição e Linguagem.

Aprovada em 27 de fevereiro de 2018, pela banca constituída pelos membros:

Prof(a). Antônio Jaeger - Orientador
UFMG

Prof(a). Rui Rothe Neves
UFMG

Prof(a). Renato Bortoloti
UFMG

Belo Horizonte, 27 de fevereiro de 2018.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Antônio Jaeger for the continuous support of study and research, for his patience and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I am particularly grateful for the assistance given by Bruna Fernanda Tolentino Moreira for providing me with invaluably constructive criticism and friendly advice during the project work.

I would also like to thank the coordinator of the sixth-grade Nalu Magalhães Salomé de Oliveira of Colégio Loyola for enabling me to collect the data needed at the school.

Finally, I wish to thank my family for their support and encouragement throughout my study.

Figure list

Second Study

Table list

First Study

Table of contents

1. General introduction

The retrieval practice has been studied and analyzed in several educational contexts. It consists in the act of retrieving, that is, recovering information that has been previously taught. The idea is to produce an effort to retrieve memories, or, in other words, to "produce an effort from within" (Roediger & Butler, 2011) as an attempt to consolidate knowledge and stimulate its retention. It is claimed that repeated retrieval, during tests, for instance, helps not only in long-term retention but also in transferring this information to another context to be used in a different manner (McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, 2011, Roediger & Butler, 2011).

The use of this strategy has been examined throughout several experiments conducted in laboratories and in real school environment, and they have shown how efficient this practice has proved to be in both artificial and actual educational contexts. The literature in retrieval practice is vast, contemplating studies with elementary school, middle school, high school, and university students, and, although several studies have reported the effectiveness of this strategy contrasted with other techniques such as rereading or highlighting, its usage is still far from been truly applied by students and teachers in classrooms (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Karpicke, 2009). Implementing this potent mnemonic strategy to achieve long-term retention has been a challenge to learners and educators specially when considering the amount and time of practice necessary to enhance the durability and efficiency of students' learning (Rawson & Dunlosky, 2011). However, regardless of the absence of this strategy in the school scenario, its efficiency is far from been contested.

The practice of retrieval is an effective learning strategy and can be incorporated to an extent of activities, such as quizzes, fill-in-the-blanks tests, multiple-choice exercises. When implemented in the classroom, students are able to recall the content of what is being learned, keeping it accessible in the future (Carpenter, 2012; Karpicke, 2012). It is important that teachers keep this in mind and start developing retrieval-based activities to help students consolidate knowledge and have access to it in the future.

Several reviews discussing the benefits of the retrieval practice have been published (Roediger & Butler, 2011; Eisenkraemer, R. E., Jaeger, A., & Stein, L. M., 2013; Balota, Duchek, & Logan, 2007; Agarwal, Bain, & Chamberlain, 2012), but none has focused exclusively on the classroom application of this practice. Thus, the first study of the current thesis is a review examining articles that show experiments which occurred in real educational environment. That is, it was

included only studies that presented empirical experiments focused on the retrieval practice with materials related to the real school content. The revision also presents a comparison among the analyzed articles, taking into account time, feedback, material, age of participants and other items.

The majority of research on retrieval practice has been conducted with college students, but just a few studies have examined its effects in children, especially those in later elementary grades (Blunt & Karpicke, 2014; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). However, research considering this school-phase is of extreme importance, since at this point students start to implement studying strategies in order to learn from what they are being taught (Karpicke, Blunt, Smith, & Karpicke, 2014). Recently, some studies have considered the effect of applying retrieval practice activities in elementary school classrooms and they have successfully confirmed that the use of tests and quizzes at this specific context can work as retrieval practice strategies to access knowledge in the long term. (Marsh, Fazio, & Goswick, 2012; Jaeger et al., 2014).

The second study presented in this thesis was conducted in an actual education context with sixth-grade children from a private elementary school in Belo Horizonte. The students performed testing and restudying tasks and their results were examined to evaluate the effectiveness of the retrieval practice. Their reading and vocabulary skills were also individually measured through the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI) and the Word/Pseudoword Reading Task developed by Salles et al (2013) in order to investigate possible associations between them and the effects of the retrieval practice. Therefore, it was possible to verify whether cognitive skills influence the effectiveness of the retrieval practice in sixth-grade students.

1.1 References

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. Educational Psychology Review, 24(3), 437-448.

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. Applied cognitive psychology, 22(7), 861-876.

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 (pp. 83-105). New York, NY: Psychology Press.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. Journal of Educational Psychology, 106(3), 849.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. Current directions in psychological science, 21(5), 279-283.

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.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. Paidéia (Ribeirão Preto), 23(56), 397-406.

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2014). Test-Enhanced Learning in Third-Grade Children. Educational Psychology, 35(4), 513-521.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. Journal of Experimental Psychology: General, 138(4), 469.

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. Current Directions in Psychological Science, 21(3), 157-163.

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: the need for guided retrieval in elementary school children. Journal of Applied Research in Memory and Cognition, 3(3), 198-206..

Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. Memory, 20(8), 899-906.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. Journal of Educational Psychology, 103(2), 399.

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

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

Salles, J. F. D., Piccolo, L. D. R., Zamo, R. D. S., & Toazza, R. (2013). Normas de desempenho em tarefa de leitura de palavras/pseudopalavras isoladas (LPI) para crianças de 1º ano a 7º ano. Estudos e Pesquisas em Psicologia, 13(2), 397-419.

1.2 Thesis structure

Following the recommendations of the Postgraduate Program in Psycology of the Federal University of Minas Gerais (UFMG), this dissertation will be presented in a scientific article format.

Study I: "Retrieval practice in the classroom: a review of applied research".

This part consists in a review of studies that investigated the use of retrieval practice as a learning strategy in actual educational contexts.

Study II: "Is retrieval practice associated with reading and vocabulary skills in sixth-grade students?"

This paper is about an empirical study that investigated the retrieval practice in sixth-grade students and verified its relationship with the learners' reading and vocabulary skills.

2. Objectives

2.1 General Objective

To analyze the retrieval practice hypothesis in educational settings.

2.2 Specific objectives

2.2.1. To associate the retrieval practice with the students' reading and vocabulary skills.

2.2.2. To verify the retrieval practice efficiency in a sixth-grade classroom.

2.2.3. To review empirical articles that investigate the retrieval practice in an educational environment.

3. First study

Retrieval practice in the classroom: a review of applied research

Tatiana Salazar Silva Pinto

Bruna Fernanda Tolentino Moreira

Danielle Veloso

Antônio Jaeger

Programa de Pós-graduação em Psicologia Universidade Federal de Minas Gerais, Brasil.
Departamento de Psicologia, FAFICH – Universidade Federal de Minas Gerais, Brasil.

Abstract

Tests have been vastly used for assessment of learning in educational contexts. Recently, however, a growing number of studies have shown that the practice of remembering previously learned information (i.e., retrieval practice) increases the long-term retention of such information; a phenomenon often termed "testing effect". The question remains, however, of whether such practice can be useful to improve learning in actual educational contexts. In the current work, we sought to address this question by conducting a review of studies that investigated the use of retrieval practice as a learning strategy in actual educational contexts. The studies selected for the analysis involved from elementary school children to college undergraduates, and the data reported by these articles show that the testing effect is fully replicated outside the lab, in educational contexts. It is important to note, however, that most reviewed studies compared retrieval practice to repeated study or to "no-activity". Further research must verify whether retrieval practice is more effective than the typical the typical activities implemented by teachers before its adoption in class can be fully recommended.

Keywords: tests, testing effect, retrieval practice, applied psychology, educational psychology.

2.2 Introduction

Retrieval practice in the classroom: a review of applied research

Tests are widely used in educational environments to assess learning of taught materials. A growing body of research, however, has shown that beyond an assessment tool, tests can also be an effective method to increase long-term retention of studied materials (Bjork, 1988; Karpicke & Roediger, 2008; Eisenkraemer, R. E., Jaeger, A., & Stein, L. M, 2013; Karpicke, 2012; Roediger & Karpicke, 2006a). Studies investigating the effects of tests on long-term memory typically involve the comparison between conditions in which participants either take tests (i.e., practice the retrieval of the studied information) or restudy a set of information. The impact of these "retrieval practices" in the long-term retention of the studied information is assessed by further tests, usually a few days or weeks later (e.g., Pashler, Rohrer, Cepeda, & Carpenter, 2007). In general, these studies show that practicing retrieval results in better long-term memory than restudying.

Although such "testing effect" has been widely replicated (see Karpicke, 2012; Eisenkraemer et al., 2013, for reviews), retrieval practices have not been commonly used in educational contexts (Roediger & Pyc, 2012). One of the reasons behind this fact might be that teachers and other education professionals are not familiar with the potential learning benefits of the testing effect. Another possible reason may be that the number of studies and experiments conducted in real educational environments remains insufficient to encourage professionals of these areas to introduce such strategy in their practices (McDaniel & Masson, 1985; Roediger & Karpicke, 2006b; Carpenter, Pashler, & Cepeda, 2009).

Despite the reasons behind the slow inclusion of retrieval practices into educational routines, a step forward in this direction may be to assess the attempts made so far to investigate empirically the application of this strategy to actual educational environments. Thus, in the current study, we review the research examining the testing effect applied in the classroom. Our main goal was to verify the effectiveness of the retrieval practice in such environments, and to examine how the success of this effect in these contexts is related to other aspects of its application, as the type of test used during retrieval practice, the presence or absence of feedback, or type of studied material. In contrast to prior reviews (Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T , 2013; Adesope, O. O., Trevisan, D. A., & Sundararajan, N. , 2017), the current study has a stronger focus on applied research, and

pursue to approach issues that are particularly relevant to the application of retrieval practices in educational contexts.

2.3 Method

In order to select the relevant articles for the present work, we searched the databases *Web of Science*, *PubMed* and *Google Scholar*, for keywords or titles containing the terms "testing effect", "test-enhanced learning", "retrieval practice", "applied", and "classroom", during the second semester of 2017. To complement the search, further searches were conducted in the reference lists of the selected materials. No restriction concerning date of publication was applied. Since the goal was to review studies applied to educational contexts only, all articles reporting laboratory studies were excluded. Thus, the inclusion criteria for the selection of studies were (1) articles should present empirical results; (2) the focus of the experiment should be on the retrieval practice; (3) studies should focus on typically developing individuals; (4) experiments should be applied to actual educational environments in the sense that (a) the to be learned materials were directly related to the content normally exposed and evaluated in particular courses/disciplines, and (b) all phases of the study were conducted in classroom environments, or, in the case of online tests, on platforms frequently used by the studied educational institutions. Studies that did not meet the remaining criteria were not included in the review.

2.4 Literature review

After an initial screening based on title and abstract, 62 articles were found from the relevant keywords. These articles were then fully examined according to the inclusion criteria, which resulted in the selection of 19 articles (see Table 1); all reporting experiments conducted in actual classrooms (see inclusion criteria above). Below, the selected articles are briefly described and discussed, focusing mainly on the results and on the experimental design, including issues as the use of a control condition, the intervals between retrieval practice/restudy and final test, and the presence or absence of feedback. To increase potential translations to actual educational settings, we provide detailed descriptions of the retrieval practice procedures adopted by each study.

Thus, a major question we raise is whether the type of retrieval practice matters. That is, is it better to conduct free recall tests than using simpler types of tests (e.g., multiple-

choice). More specifically, what is the current evidence in favor of the application of each type of retrieval practice? This question is approached below.

Does free recall increase learning in the classroom?

Free-recall has been shown to produce strong testing effects in laboratory settings (e.g., Roediger & Karpicke, 2006a). To verify the effectiveness of such type of test in the classroom is rather challenging, however. The materials typically used in the classroom are presented in a less controlled manner than it is done in laboratory settings. That is, while in laboratory settings participants usually encode lists of words, often controlled for concreteness, frequency, number of letters, which are presented individually and for a controlled amount of time; in the classroom, materials are often presented in lectures, texts, discussions, among others, which are conducted in groups including several students. In spite of these challenges, a few studies investigated whether conducting free-recall tests could be an effective strategy to improve learning in the classroom.

One such study was reported by Dobson and Linderholm (2015), and involved two phases. In phase 1, students of an anatomy and physiology course studied three passages describing structures and concepts of (1) cardiac electrophysiology, (2) ventilation, and (3) endocrinology (contents to be studied in future semesters). Each passage had a little more than 600 words, and the study of each passage was conducted in different conditions. That is, each student was randomly assigned to (a) read one of the passages three times in a row (R-R-R), (b) read the second passage and then reread it while taking notes (R-R + N), and (c) read the third passage, complete a free recall task, and then read it again (R-T-R). The free recall task consisted of writing down in a blank sheet as many concepts and definitions from the passage as possible. Immediately after this, and once again one week later, participants performed multiple-choice tests on the three studied contents. In the first test (immediate), no significant difference in performance between the R-T-R and R-R+N groups was found, although both groups performed better than the R-R-R group. In the delayed test (1 week later), however, the R-T-R group showed greater performance than the other two groups. In the second phase of the study, the results of the first phase were presented to the participants, and they were encouraged to use the superior R-T-R strategy to get prepared for the remaining course exams. The use of such strategy significantly increased performance on the course exams, suggesting that when students adopt self-testing in their preparation for exams, learning can be significantly improved. Importantly, this was the only free-recall study in which the impact of tests in the actual course exams was assessed.

Another study examining whether free-recall tests can enhance learning in classroom settings was reported by Lipko-Speed, Dunlosky and Rawson (2014). This study involved 5th graders and focused on the science contents that are typically studied in this grade, namely, light and sound (experiment 1), and geography (experiment 2). This study comprised four sessions. The first and second sessions were administered with an interval of 48 hours between them (on Monday and Wednesday of the same week). The third and fourth sessions were administered one week later, also with a 2-day interval between them (on Monday and Wednesday of the following week). In the first session, all participants studied 20 key terms and their definitions (e.g., What is sound? Form of energy that you can hear that travels through matter as waves). Five of the 20 key terms were then assigned as controls, and not presented again until the fourth session (final test). During the second and third sessions, the remaining 15 key terms were assigned to three conditions (five to each condition). In the "test" condition, participants were required to type the definitions of each key term (What is sound? _____); in the "test-plus-feedback" condition, participants received feedback after typing the definition of each key term, and in the "study" condition, each definition was presented again, and participants studied them in a self-paced manner. Finally, in the fourth session (final test), students were asked to type the concept of all 20 key terms. Items assigned to the test-plus-feedback condition were better remembered in the final test than items that were just tested or restudied. Surprisingly, however, the contrast between the recall of tested items (without feedback) vs. restudied items in the final test was not significant, suggesting that feedback was essential to increase learning.

Finally, in the study reported by Carpenter, S. K., Lund, T. J., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016), introductory biology undergraduates completed in-class exercises that were followed by the recall or the copy of the studied terms' definitions (e.g., Polar body: a cell produced by asymmetric cell divisions during meiosis). After their responses, a corrective feedback was given. Five days later, all students completed an unannounced quiz assessing the information learned from the exercises. Their results showed that only high-performing students (as assessed by prior course exams) were benefited by recalling the definitions of the terms, while low- and middle-performing students were more benefited by copying the definitions than recalling them.

Overall, these studies employed robust experimental designs, and they all included restudy as a control condition (plus taking notes in Dobson & Linderholm, 2015). Unfortunately, however, they did not verify whether recall produced long-term learning, since they only tested memory for a few days after retrieval practice. Furthermore, all three studies provided feedback in at least one testing condition, and when feedback was not absent, the

benefits of retrieval practice were absent (Lipko-Speed et al. 2014).

Although the results of these studies are mostly favorable to the use of free recall as a retrieval practice in the classroom, in one of them (Carpenter et al. 2016) no positive effect of retrieval practice was found (i.e., recall versus copy). That is, such effect was only present for high-performing individuals, and it was actually negative for the remaining students. Also, only one study analyzed the influence of recall in the actual exams of the course (Dobson & Linderholm, 2015), while the remaining studies relied only on tests conducted by the researchers. Finally, the presentation of the to-be-learned materials were not necessarily analogous to the manner materials are presented during actual classes. Students actually studied the materials individually by reading a text (Dobson & Linderholm, 2015), or by reading key terms followed by their definitions (Lipko-Speed et al. 2014). Would the positive effects of retrieval practice persist if the studied materials were presented in a power point lecture, or class discussion?

Taking into account the above data, is it possible to conclude that free recall increase learning in the classroom? Although we can speculate based on these results that free recall tests accompanied by feedback show some benefits over coarser learning strategies in the classroom (e.g., restudy the same material multiple times), substantial work should be done before strong recommendations in favor of the use of such practices in the classroom can be safely made. The number of applied studies found is very small, the intervals between retrieval practice and final test adopted by them are short, only one study recruited children as participants (Lipko-Speed et al. 2014), and in one study retrieval practice was advantageous for high-performing students only (Carpenter et al. 2016). Thus, even though a large number of laboratory studies posits free recall as the best retrieval practice to enhance long-term retention (Dunlosky et al. 2013), it remains an open question whether this practice is equally powerful in applied contexts.

Do short-answer tests increase learning in the classroom?

Several studies show that cued-recall tests are highly effective in eliciting testing effects (see Eisenkraemer et al. 2013, for a review). In the classroom, cued-recall tests are often implemented in the form of short-answer tests. Below, we review studies that examine whether (and how) such type of test can be useful to enhance learning in the classroom.

Leeming (2002) assessed the use of short-answer tests to increase learning during two regular classes in an introductory psychology course (i.e., Introductory Psychology, and Learning and Memory), and used short-answer and short-essay tests to explore it. While half

students performed only the usual three examinations during the semester, the remaining students performed daily tests during the whole course. The daily tests lasted from 10 to 15 minutes and consisted of two short-essay questions taken from the book adopted in the course, and five short-answer questions based on the content of the texts and the content of the lessons. Immediately after each daily test, the researcher commented and corrected the inappropriate responses (i.e., gave feedback). All participants took a final test about 6 weeks after classes were finished. The final test contained short-essays, multiple-choice, and fill-in-the-gaps questions about the issues covered by the daily tests. The results showed that the students who performed daily tests often had better performance on the final test than the students who did not take them. These results, however, do not make it clear if students performing tests in all classes exhibited a greater delayed performance due to becoming more involved in the learning process (because they were studying harder and paying more attention to classes, since they knew they would be tested), or due to content retention as a result of the testing effect itself. Despite of this, the data suggests that the practice of frequent and systematic examinations can be a strong strategy to enhance learning in the classroom.

Lyle and Crawford (2011), following the same line, assessed the effect of answering short-answer questions in the end of each class of a statistics course. The authors of the study taught this course to psychology students in two consecutive years. Tests were administered only at the second year, and performance at the first year served just for comparison. The tests consisted of short-answer questions projected on a whiteboard during the last ten to fifteen minutes of each class. Students were required to write down and submit their responses before leaving class. At the beginning of each class, the correct answers (i.e., feedback) for the questions from the previous class were projected on the whiteboard (and posted on the course website), and participants were free to review their responses and make further questions before new materials began to be taught. In order to motivate students to perform the tests, they counted for 8% of the course grade. Performance of students in the comparison group (without end of class tests) and students in the test group were assessed by four multiple-choice tests distributed throughout the semester. The exam questions did not use the same wording as the test questions, but their main contents were the same. In three out of the four examinations of the course, participants who did take tests performed significantly better than participants who did not. In addition, it was also found that the students' scores in the end-of-the-class tests were positively correlated with the grades obtained in the exams.

To investigate whether short-answer tests followed by feedback increase the long-term retention (six months) of materials studied in medical courses, Larsen et al. (2009) recruited

medical residents (pediatrics and emergency) who attended lectures about myasthenia gravis and epilepsy. Half the residents were tested on epilepsy and restudied myasthenia gravis, while the other half was tested on myasthenia gravis and restudied epilepsy. Test and restudy sessions were held both immediately after the lectures, and after a two weeks interval. They were followed by corrective feedback. In a final test, administered 6 months after the second restudy/test session, materials tested by short-answer questions were significantly more remembered than restudied materials.

The potential benefits of short-answer tests (and multiple-choice) for learning were also examined by Burdo and O'Dwyer (2015). In their study, psychology students enrolled in a physiology class were assigned to three groups. Group 1 attended to twelve encounters during the semester to review the studied materials using a "mapping concepts method" (Novak, 2010). Group 2 attended to twelve encounters in which the retrieval of the studied materials was practiced. To perform the retrieval practice, students were divided into small groups and had to elaborate short-answer and multiple-choice questions using the available written materials as much as they felt necessary. They then exchanged the questions with each other so that other members of the group could answer the questions (without access to written materials). The questions then returned to their authors to be corrected, and then correct and incorrect answers were discussed among the group members (i.e., corrective feedback). Group 3 (control group) did not attend to such extra-class meetings as the other two groups. Learning in each of these conditions was assessed through five tests placed during the semester (4 tests of materials studied in specific units and 1 cumulative final test), and the course's final grade. The analyses showed that in the units' tests and in the final grade, students in the practice group (i.e., group 2) had a numerically (non-significant) higher performance than the other two groups, but only in the second unit test this difference reached statistical significance. Furthermore, even though no differences in performance among the groups were found in the final exam, the control group had a numerically greater performance than the other two groups in this test.

Finally, Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009) examined the potential benefits of short-answer questions for the learning of history contents by 8th grade students. The questions comprised materials exposed to students during class, in discussions, notes, reading assignments, and handouts. The tests (and restudy) were administered 1-week ("immediate review" group) or 4 months ("late review" group) after the first exposure of the students to the to-be-learned materials. The tests were in the form of simple history questions (Who assassinated President Abraham Lincoln?) to which participants should give a short response (John Wilkes Booth)

followed by corrective feedback (i.e., the correct response). A final test administered nine months after the students had completed the review, showed increased long-term retention for tested items relative to restudied or studied only items.

Overall, the design employed in the short-answer studies involved restudy or regular class as control condition, involved various intervals between retrieval practice and final test (i.e., from a few weeks to 9 months), and included corrective feedback. Only one study involved children (Carpenter et al. 2009), and all the remaining studies involved undergraduates or medical students.

One study showed only weak evidence in favor of short-answer tests (Burdo & O'Dwyer, 2015). Perhaps the absence of reliable effects in this study is related to how the retrieval practice was implemented. Note that in that study, students had to elaborate questions to their peers, correct them afterwards, and then discuss the responses with the group. Curiously, however, these were extra-class activities. Thus, a mnemonic advantage would be expected due to greater exposure to the studied contents.

Two of the reviewed studies implemented the retrieval practice as a daily routine in the classroom (Leeming, 2002; Lyle & Crawford, 2011) and showed encouraging results. The retrieval practices adopted in these studies are relatively easy to implement in most undergraduate courses, and do not take too much time away from other activities administered in class. It is unknown, however, whether using short-answer questions in such a systematic manner would promote learning for younger students (e.g., mid-school children), or whether some adjustments in the way the practice was applied would be necessary. These important issues should be addressed by futures research.

Does simply asking students to fill in gaps increase learning in the classroom?

In general, short-answer questions seem to be a promising strategy to enhance learning in the classroom. Does a slightly simpler type of "short-answer" test elicit similar results? In the studies reviewed below, students were required to simple fill-in-the-gaps of texts. As for short-answer tests, such task is analogous to cued recall tests, a type of test shown to be effective in eliciting testing effects (Kang et al. 2007).

The first study was reported by Cranney, Ahn, McKinnon, Morris and Watts (2009), and assessed whether the potential benefits of "fill-in-the-gaps" tests along with multiple-choice questions in two experiments in which psychology students served as subjects. The first experiment was divided in three phases: (1) initial learning, (2) review and (3) final

memory test. In the initial learning phase (1), all students watched a video about Psychobiology, which consisted in an introduction on brain signaling and brain structures. In the review phase (2), students were assigned to one of the following groups: a) groups of 4 to 5 people performed a quiz about the video collaboratively, b) students completed the same quiz individually, c) students read a summary of the video and highlighted the most important parts and had 2 minutes to ask any questions, and d) students in the no-activity condition did not re-engage with the video information. In the conditions including responding to quizzes, students had to answer multiple-choice and fill-in-the-gaps questions about the film (e.g., the hippocampus is part of the _____ system [*limbic*]), and all their answers were followed by corrective feedback. In the final memory test (3), one week after the review phase, participants who performed quizzes in groups showed greater performance than participants who performed quizzes individually (and from those who performed no quiz). In experiment 2, the group quiz condition was excluded, and participants in the individual quiz condition showed greater performance than participants in the restudy and participants in the no-activity condition. Unfortunately, no comparisons between the results for fill-in-the-gaps and multiple-choice tests were made, thus it is not possible to infer which test was more important for the results.

The second study assessing fill-in-the-gap tests (Vojdanoska, Cranney & Newell, 2010) involved three phases. In the first phase participants attended to a 10-minutes long power point presentation about adult development. The presentation included short videos and contained at least 24 relevant items of information. In the second phase, which was immediately after the first, participants responded in groups or individually to a test containing 16 fill-in-the-blanks questions (out of the 24 relevant items). Importantly, half of these questions were followed by feedback. In the third phase, which was held a week later, participants responded to a final test, which included the 16 questions presented in the second phase and the remaining 8 untested questions. A further group of participants (control group) skipped the first and second phases, and performed just the final test. The data showed that students who were subjected to the first and second phases showed greater performance than the control group in the final test. More importantly, participants showed greater performance for questions that had been tested in comparison to untested questions, and greater performance for questions followed by feedback than for questions not followed by feedback. Conducting tests individually or in groups resulted in indistinguishable performance at the final test.

In the final fill-in-the-gaps experiment (Jaeger Jaeger, A., Eisenkraemer, R. E., & Stein, L. M.. 2015), third-grade children read

twice a 321-word encyclopedic text about the Sun, which contained 20 key terms. After reading the text twice, students either read the text twice again, or received the text with the 20 key terms missing, and had to write them down in the appropriate gap (e.g., The superficial layer of the sun is called _____ [*photosphere*]). Responses were not followed by feedback. In the final multiple-choice test administered seven days later, students in the "fill-in-the-gaps" condition had a very superior performance. Also, measures of individual differences in IQ and reading skills conducted in this study suggest that practicing retrieval can benefit all children within the normal range of these abilities.

The fill-in-the-gaps studies described above included a restudy control condition, except for the study reported by Vojdanoska et al. (2010), which compared the test conditions to a "no-activity" control group, and to a condition in which items were only studied, but not restudied. The final test in all these studies were administered one week after retrieval practice. Feedback was present in one study (Cranney et al. 2009), was absent in one study (Jaeger et al. 2015), and was present for half the test items in one study (Vojdanoska et al. 2010). Although retrieval practice resulted in positive effects in all cases, Vojdanoska et al. (2010) showed that feedback increases this effect when available.

Thus, overall, a simple type of short-answer test (i.e., fill-in-the-gaps) yielded strong testing effects in actual educational environments for questions about a video presentation (Cranney et al. 2009), questions about a power point presentation, with or without feedback (Vojdanoska et al. 2010), and questions about an encyclopedic text read by third grade children. Thus, although the number of studies assessing the application of fill-in-the-gap tests in actual classrooms remains small, this simple strategy seems promising, especially taking into account that it is easy to apply, is not time consuming, and is relatively easy to correct and to give feedback.

Do multiple-choice questions increase learning in the classroom?

Multiple-choice questions are frequently used to evaluate learning, since they are easy to administer and easy to correct. The question at issue here is whether multiple-choice tests can be a useful strategy to enhance learning in the classroom. The answer for this question is not obvious, since several laboratory experiments suggest that multiple-choice tests are not as advantageous as cued recall and free recall for long-term retention (Kang, S. H. K., McDermott, K. B., Roediger, H. L. , 2007). An advantage of this type of test, however, is that it can be performed at home, leaving more time available in class for further teaching activities (Kornell, N., Rabelo, V. C., & Klein, P. C. , 2012). That is, teachers can add such practice without having to sacrifice other activities they usually implement in class.

Thus, in a study to investigate whether multiple-choice tests can improve retention of science contents, McDaniel, Agarwal, Huelser, McDermott and Roediger (2011) conducted three experiments in which 8th grade students performed three consecutive multiple choice "quizzes", followed by corrective feedback. The first quiz was applied before class (students had been instructed to read a text about the lesson at home), the second quiz was applied immediately after class, and the third quiz was applied 24 hours before each unit exam (20 days after the first quiz, on average). Overall, participants who were quizzed showed increased performance in the unit exams, in the end-of-semester, and end-of-year exams in comparison to participants who were not quizzed. Interestingly, quizzes administered 24 hours before the unit exams increased performance not only in the unit exams, but in the end-of-semester exams as well. This effect was persevered even when earlier quizzes were absent, suggesting that conducting multiple-choice tests before exams can be an effective tool to improve both exam performance and long-term retention. Importantly, these findings were replicated in a follow-up study (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013) wherein the format of the questions were changed on each of the three quizzes. That is, some questions focused on applications of the concepts, others on definitions of the concepts, and so on.

The potential learning benefits of multiple-choice questions in the classroom was also examined by Roediger, Agarwal, McDaniel & McDermott (2011). The authors conducted three experiments in 6th grade social studies classes. In the first experiment, a series of multiple-choice quizzes were administered: a) pre-test quizzes, administered immediately before class; b) post-test quizzes, administered immediately after class; c) review tests, administered a few days after class; d) chapter exams, administered about 2 days after each review test; and e) a late exam, which was administered at the end of the semester (about 1 or 2 months after study). The quizzes (pre-test, post-test, and review test) covered all the taught materials and consisted of 4-alternative multiple-choice tests. The chapter exams consisted of an initial free recall test of all contents of the chapter, followed by multiple-choice questions about quizzed and unquizzed items. Students exhibited greater performance at the chapter exams for quizzed than for unquizzed materials, and such enhanced performance was replicated in the later exams. In the second experiment, a control group that just read the quiz questions was included, and the enhanced retention of quizzed materials found in experiment 1 was replicated in the chapter exam, and partially replicated in the late (end-of-semester) exam. Finally, in the third experiment, students were encouraged to use a website to test their learning. The website could be accessed from the students' homes, and offered different kinds

of tests, including games requiring responses to relevant questions. This experiment differed from experiments 1 and 2 by keeping only pre-test quizzes (i.e., post-test and review tests were removed). In total, eight pre-test quizzes were given throughout the year. In the chapter exams (multiple-choice and short-answer) and in the end-of-semester exam (multiple-choice, administered 1-3 months after the target materials were initially studied), the effect of test persisted for pre-tested items, despite of whether participants used the website or not. Overall, this study shows that taking multiple-choice tests can benefit later performance in multiple-choice and short-essay tests, suggesting that such practice is flexible in the sense that practicing retrieval in one type of test can enhance performance in a different type of test.

The four remaining multiple-choice studies were conducted on psychology (Balch, 1998; Batsell, Perry, Hanley & Hostetter, 2016; Daniel & Broida, 2004) and physiology classes (Kibble, 2007). In the study conducted by Balch (1998), students in an introductory psychology class were divided in two groups, the "practice-exam" group and the "review-exam" group. Participants in the practice-exam group received a multiple-choice test with questions similar in format and content, but not identical, to the questions they would later encounter in a final exam. Responses to such questions were followed by corrective feedback. Participants in the review-exam group were asked to read the same questions, along with their answers, and to rate them in a 4-point scale regarding their expectancy of finding these questions again in the final test. In the final test, a week later, students from the practice-exam group exhibited superior performance than participants from the review-exam group.

Batsell, et al. (2016) assigned students from an introductory psychology course to a control group, which received daily readings; and to a quiz group, which received daily readings and completed daily multiple-choice quizzes on the readings. Students from both groups performed three exams containing multiple-choice questions throughout the semester. The questions of the exams were divided in three types: identical questions (the actual quizzed questions), new questions, and similar questions. The similar questions covered topics that had been quizzed, but were substantially different from the questions of the quizzes applied before. The use of daily quizzes improved performance on identical, similar, and new questions compared to the no-activity ("control") condition. Notably, feedback was not given in this study, and the testing effect was assessed only for materials studied in the textbook (i.e., not studied in-class).

In the study reported by Daniel and Broida (2004), psychology classes were assigned to one of three conditions: no quiz, in-class quiz, and web-based quiz. In the in-class quiz condition, students responded to quizzes (i.e., 10 chapter-based questions) during the first

fifteen minutes of each class. In the web-based quiz condition, the same quizzes were available online for the students during the 24 hours preceding each class (for 15 minutes once they get started). In both conditions, students received immediate feedback of their performance on each quiz. In the no quiz condition, students did not have to complete any kind of quiz, nor did they have access to them. Performance was assessed in four multiple-choice and/or short-answer tests throughout the semester. As expected, the in-class quiz group performed significantly better than the no-quiz group. However, the performance of the web-based quiz group did not differ significantly from the no-quiz group. The authors were later informed that students were resorting to various strategies to cheat while taking the online quiz. Thus, to avoid this, some changes were made in the online quiz. Specifically, the 10 questions of each quiz were randomly selected from a pool of 100 questions, and the time to complete each quiz was reduced from 15 to 7 minutes. After those changes, students took two more examinations and both experimental groups performance (i.e., web-based quiz and in-class quiz) were equivalent, and significantly greater than the performance of the no quiz group.

Finally, in the study reported by Kibble (2007), online quizzes were available to help students from five different classes of physiology (approximately 350 students per class) to get prepared for their usual middle- and end-of-semester exams. Two quizzes were available before each exam, and corrective feedback was given after the second quiz. For each of the five classes, the quizzes were offered within different reward models. In model 1, no incentive was given for the completion of the quizzes; in model 2, 0.5% of the grade was given for those who accessed the quizzes; in model 3, 1% of the grade was given for those who responded correctly to 30% of the quizzes or more; in model 4, 1% of the grade was offered according to the scores obtained in the best of two attempts to complete the quizzes; and in model 5, 2% of the grade was given using the same criteria of model 4. In the class of model 1, as no incentive was offered for the completion of the quizzes, only 52% of the students performed them. Those who completed at least one quiz in this group, performed significantly better on the tests than those who did not perform any quiz. In the other classes, incentives increased dramatically the access to quizzes, resulting in a very small number of students who did not take any quiz. Interestingly, greater incentives led some students to take quizzes incorrectly, using previously written notes or the help of other people. These students, who achieved a score close to 100% in their first attempt, did not achieve a grade as good as the students who used the two attempts to complete the quiz.

Overall, from the studies assessing exclusively the use of multiple-choice tests during

retrieval practice, only two recruited children as participants (McDaniel et al. 2011; 2013; Roediger et al. 2011). The remaining studies assessed such approach on undergraduate or medical school students (Balch, 1998; Daniel & Broida, 2004; Batsell et al. 2016; Kibble, 2007). The interval between retrieval practice and final test ranged from one week (Balch, 1998) to a few months (e.g., Roediger et al. 2011). Only two studies did not include feedback (Batsell et al. 2016; Haynie, 1997), but retrieval practice yielded positive effects in both cases. The results from all these studies were in favor of the use of multiple-choice tests in the classroom.

Taking into account these results and the fact that multiple-choice tests are easy to apply, easy to grade, and easy to give feedback, such type of test emerges as a recommendable practice to improve learning. Notably, however, further research is still needed to verify the usefulness of such practices for further materials, since in all cases reported here, they were applied to science and technology studies.

Interestingly, Kibble (2007) final tests were conducted several weeks after study, but quizzes could be performed immediately before the class' usual exams. It is interesting to note that this study highlights some of the difficulties of using online quizzes as retrieval practice for materials studied in the classroom. While students access more the online quizzes when reward is provided, and it results in greater learning, care should be taken in order to prevent students from not following the rules when using quizzes (see also Daniel & Broida, 2004).

Is there a preferable type of retrieval practice to be used in class?

The studies reviewed above did not analyzed separately the effects of different types of retrieval practice on learning. We found only two studies that conducted analysis directly comparing multiple-choice to short-answer tests.

In the study reported by Ramraje and Sable (2011), medical school students enrolled in Pathology classes were assigned to one of three treatments: (1) multiple-choice tests administered in the end of the class, (2) short-answer tests administered in the end of the class (3), absence of tests in the end of the class. After three weeks, all students completed an unexpected test comprising multiple-choice and short-answer questions. As expected, both multiple-choice and short-answer tests resulted in greater performance relative to no test in the later tests. Surprisingly, however, in the final test participants who performed multiple-choice tests showed greater performance than participants who performed short-answer tests.

McDermott et al (2014) also examined whether short-answer and multiple-choice questions were differentially beneficial to enhance retention in the classroom. Their study

comprised a series of experiments involving seventh-grade science and high school history contents. In all experiments, students took intermittent quizzes (short-answer or multiple-choice, both with corrective feedback) and performance on unit-exams and end-of-semester exams were analyzed. The quizzes were administered immediately before the materials were taught (prelesson quiz), immediately after the materials were taught (postlesson quiz), and one day before the unit exam (review quiz), which took place a few days after the postlesson test. Both multiple-choice and short-answer questions were projected on a screen in front of the classroom, and the research assistant read the questions aloud to the students. The students responded within a limited amount of time by using clickers or writing down their responses in a paper. The restudy condition consisted in projecting the question with its most appropriate answer. The results were clearly favorable to practicing retrieval relative to restudying the materials. More importantly, short-answer and multiple-choice questions were equally effective in enhancing retention of studied materials.

Thus, even though both studies were favorable to the use of multiple-choice and short-answer questions relative to restudy or no-activity, multiple-choice tests showed some advantage in one of the studies (Ramraje & Sable, 2011) and was equivalent to short-answer in the other (McDermott et al. 2014). Taking into account the positive results encountered in the studies using exclusively multiple-choice questions (see section above), and the application and grading advantages of this type of practice, such strategy arises as a highly promising approach to be used in class. Especially considering that multiple-choice tests demand considerably less classroom time than the remaining tests.

2.5 Integrative conclusion

The reviewed studies suggest that promoting retrieval practice at class is a promising approach to enhance the long-term retention of studied materials. Positive effects were found for different intervals between retrieval practice and final tests, as well as for several types of materials and encoding tasks (see Table 1). Furthermore, feedback was not essential for these benefits to emerge, although it was considerably helpful. Even though such results are favorable to the use of retrieval practices in the classroom, some important issues should be considered regarding the reviewed data. First, the practice of retrieval was mostly compared with the restudy of the taught materials, or with no activity at all. Second, the applied data do not seem to mirror the ubiquitous laboratory finding that more effortful retrieval practices

result in better long-term retention than less effortful ones (e.g., Kang et al. 2007).

The first issue is perhaps the most challenging for the translation of the reviewed data to actual class environments. As can be seen in Table 1, 10 studies compared retrieval practice to restudy, 6 compared retrieval practice to "no activity", one compared retrieval practice to a "mapping concepts" method about the studied materials (and found no advantage for retrieval practice over this technique), and two compared retrieval practice to the last 10 or 15 minutes of regular class. Thus, while retrieval practice was compared to "nothing" in several studies, the control condition most often used was the restudy of the same material in the same way it was initially studied (although some studies slightly departed from this, see Table 1). This (restudy) is a considerably weak learning strategy (e.g., Callender & McDaniel, 2009), which is rarely used in class because even when materials are repeatedly presented to students, the usual classroom activities (i.e., lectures, discussions) makes such exact repetition activity difficult to conduct.

Critically, teachers have a limited amount of time in class, thus whenever they choose to include a new activity to their routines, they usually have to sacrifice other activities (See Kornell et al. 2012, for a discussion on this issue). The question then is whether the activities teachers will have to sacrifice to add retrieval practice to their classes are as weak as restudy or "no-activity" in promoting long-term learning. To approach this problem, future research should compare retrieval practice with activities that teachers actually do at class. The only two studies that approached this (Leeming, 2002; Lyle et al. 2011) showed promising results (although note that the last 15 minutes of a lecture is not necessarily a strong control condition). A fruitful future approach then could be to first survey the activities teachers actually do in class (through questionnaires or observations), and then to compare these activities to different types of retrieval practices.

The second issue raised above, that is, the absence of evidence suggesting that more effortful retrieval yields greater learning, has profound implications for the design of retrieval practices for actual classes. Note that more effortful retrieval practices (e.g., free recall) take more class time, are harder to grade, and harder to give feedback to. Interestingly, in the reviewed data, multiple-choice tests, which are the easier retrieval practices to be implemented in classrooms, showed results that were equivalent or better than more effortful tasks (e.g., cued recall tasks, McDermott et al. 2014; Ramraje & Sable, 2011). Considering that multiple-choice tasks are substantially less time demanding, and can be conducted at the students' homes, it emerges as a promising possibility, which does not suffer from the class time issues discussed in the above paragraphs.

Turning to feedback, does it matter according to the current findings? In most of the reviewed studies, participants received feedback after test, which varied only in the way it was presented (e.g., projected on the board, available online, etc.). The experimenters' choice for this practice can be justified by previous demonstrations that feedback can boost the testing effect (e.g., Butler & Roediger, 2008). However, in the few studies in which feedback was absent, the testing effect was equally replicated (see Ramraje and Sable, 2011). In addition, in the study by Vojdanoska et al., (2010), which directly compared the effectiveness of using questions with and without feedback, no significant difference was found between these conditions. Thus, although further research will be needed to verify whether feedback can significantly boost the testing effect in diverse classroom environments, the currently available literature suggests that feedback is perhaps not a crucial ingredient for the successful implementation of teaching strategies based on retrieval practice.

Overall, the studies reviewed here reveal that the testing effect is not just a laboratory phenomenon; it can be fully replicated in real educational contexts. This is shown here for different educational contexts, different contents (e.g., physiology, social sciences, etc), different types of tests, different intervals between tests, and different populations. The reviewed research shows that the retrieval practice can be implemented directly in the class format, and this implementation can take many formats; from small tests administered during or at the end of each class (e.g., Lyle & Crawford, 2011), to simply teaching students about the efficacy of the testing effect (Dobson and Linderholm, 2015). Although the current evidence is overall favorable to the use of retrieval practice at class, an important issue is related to the activities that will be sacrificed to open space for retrieval practice. A challenge for researchers concerned about educational psychology is to discover whether retrieval practices are more advantageous for learning than activities teachers have been typically doing at class. That is, is practicing retrieval better for learning than a lively discussion about the learning materials, or than an instigating lecture? Based on the reviewed data, the only recommendation for teachers that can be done now is to substitute simple restudy activities (or "no-activity") for retrieval practice. Further research will be needed before the substitution of other teaching activities for retrieval practice can be safely recommended.

2.6 References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. **Review of Educational Research, 87,** 659-701.

Balch, W. R. (1998). Practice versus review exams and final exam performance. **Teaching of Psychology, 25**, 181-184.

Batsell, W. R.; Perry, J. L.; Hanley, E; Hostetter, A. B. (2016). Ecological Validity of the Testing Effect. **Teaching of Psychology, 44,** 18-23.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory II (pp. 396-401). London: Wiley.

Burdo, J., & O'Dwyer, L. (2015). The effectiveness of concept mapping and retrieval practice as learning strategies in an undergraduate physiology course. **Advances in physiology education**, **39**, 335-340.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. **Memory & Cognition, 36,** 604-616.

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. **Contemporary Educational Psychology, 34,** 30-41.

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., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. **Applied Cognitive Psychology, 23,** 760-771.

Carpenter, S. K., Lund, T. J., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. **Educational Psychology Review**, **28**(2), 353-375.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. **European Journal of Cognitive Psychology**, **21**, 919-940.

Daniel, D. B., & Broida, S. (2004). Using web based quizzing to improve exam performance: Lessons learned. **Teaching of Psychology 31**, 207-208.

Dobson, J. L., & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. **Advances in Health Sciences Education**, **20**, 149-161.

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,** 4-58.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. **Paideia, 23,** 397-406.

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2015). Test-enhanced learning in third-grade children. **Educational Psychology, 35,** 513-521.

Kang, S. H. K., 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,** 528-558.

Karpicke, J. D. (2012). Retrieval-based learning: active retrieval promotes meaningful learning. **Current Directions in Psychological Science, 21,** 157–163.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. **Science, 319,** 966-968.

Kibble J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. **Advances in Physiology Education, 31,** 253–260.

Kornell, N., Rabelo, V. C., & Klein, P. C. (2012). Tests enhance learning – Compared to what? Journal of Applied Research in Memory and Cognition, 1, 257–259.

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: A randomized controlled trial. **Medical Education, 43,** 1174–1181.

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. **Teaching of Psychology, 29**, 210-212.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? **Journal of Applied Research in Memory and Cognition, 3,** 171-176.

Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. **Teaching of Psychology**, **38**, 94-97.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. **Journal of Experimental Psychology: Learning, Memory and Cognition, 11,** 371 - 385.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., Roediger, H. L. (2011). Test-enhanced learning in a middle school classroom: the effects of quiz frequency and placement. **Journal of Educational Psychology, 103,** 399–414.

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**, 360-372.

McDermott, K. B., Agarwal, P. K., D'antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. **Journal of Experimental Psychology: Applied**, **20**(1), 3.

Novak, J. D. (2010). Learning, creating, and using knowledge: concept maps as facilitative tools in schools and corporations. **Journal of e-Learning and Knowledge Society, 6,** 21-30.

Pashler, H., Rohrer, D., Cepeda, N. J., Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: choices and consequences. **Psychonomic Bulletin & Review, 14,** 187-193.

Ramraje, S. N., & Sable, P. L. (2011). Comparison of the effect of post-instruction multiple-choice and short-answer tests on delayed retention learning. **The Australasian medical journal**, **4**, 332-339.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. **Perspectives on Psychological Science, 1,** 181–210. **Psychological Science, 17,** 249–255.

Roediger, H. L., Agarwal, P. K.; McDaniel, M. A.; McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. **Journal of Experimental Psychology: Applied, 17,** 382-395.

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: applying cognitive psychology to enhance educational practice. **Journal of Applied Research in Memory and Cognition, 1,** 242-248.

Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. **Applied Cognitive Psychology**, **24**, 1183-119

3.6 Appendices

Table 1. Main procedures used in the reviewed articles.

References	Participants/course	Actual exams included.	Study task	Retrieval practice test	Control condition	Feedback
Dobson *et al.*, 2015	Anatomy and physiology students.	Yes	To read a 600-words text	Free recall	Restudy and taking notes	Present
Lipko-Speed *et al.*, 2014	Tests of science concepts administered to 5th graders	No	Read terms and their definitions	Free recall	Restudy and no activity	Present
Carpenter *et al*., 2016	Undergraduate introductory biology course students	No	Assigned readings	Free recall	Copy written materials	Present
Leeming, (2002)	Psychology students.	Yes	Lectures	Short-essay and short-answer	Lecture	Present
Lyle *et al.*, 2011	Psychology students	Yes	Lectures	Short-answer	Lecture	Present
Larsen *et al.*, 2009	Pediatrics and emergency medicine residents	No	Lectures	Short-answer	Restudy	Present
Burdo *et al.*, 2015	Psychology students	Yes	Lectures	Multiple-choice and short-answer	Mapping concepts method or no activity	Present
Carpenter *et al.*, 2009	8^{th} grade history students.	No	Discussions, notes, readings and handouts	Short-answer	Restudy	Present

Cranney *et al.*, 2009	Psychology students	No	Video presentation	Fill-in-the-gaps + Multiple-choice	Restudy (Read and highlight a summary of the video) or no activity	Present
Vojdanoska *et al.*, 2010	First year psychology students	No	Power-point presentation	Fill-in-the-gaps	No activity	Present
Eisenkraemer *et al.*, 2013	Third-grade students	No	Read a 321-words text	Fill-in-the-gaps	Restudy	Absent
McDaniel *et al.*, 2011	Science questions applied to 8^{th} students	Yes	Assigned readings and lectures	Multiple-choice	No activity	Present
Roediger III *et al.*, 2011	Social studies questions applied to 6th graders.	Yes	Lectures	Multiple-choice	Restudy and no activity	Present
Balch, (1998)	Introductory psychology students	Yes	Lectures	Multiple-choice	Restudy and rate expectancy of finding question in final exam	Present
Batsell *et al.*, 2016	Psychology students	Yes	Assigned readings	Multiple-choice	No activity	Absent
Daniel & Broida, 2004	Psychology students	Yes	Lectures and assigned readings	Multiple-choice	No activity	Present
Kibble (2007)	Medical students	Yes	Lectures	Multiple-choice	No activity	Present
Ramraje *et al.*, 2011	Medical students	No	Lectures	Multiple-choice and short-answer	No activity	Absent
McDermott *et al.*, 2014	7th grade science and high school history	Yes	Lectures	Multiple-choice and short answer	Restudy	Present

4. Second study

Is retrieval practice associated with reading and vocabulary skills in sixth-grade students?

Tatiana Salazar Silva Pinto

Programa de Pós-graduação em Psicologia, Universidade Federal de Minas Gerais, Brasil. Departamento de Psicologia, FAFICH – Universidade Federal de Minas Gerais, Brasil.

Abstract

Recent research in cognitive psychology have shown that taking a test on certain information is better to promote memory retention than just restudying or rereading that information. This phenomenon, termed the testing effect, has proved to be a relevant educational tool since it enhances memory performance collaborating to knowledge consolidation. In this article, we report an experiment in which the participants were sixth-grade children. The students initially read a brief encyclopedic text twice. As soon as they finished, they did a distractor activity and then performed a fill-in-the-blanks test with ten words missing and restudied other ten which were in bold. After a seven-day interval, a fill-in-the-blanks and a multiple-choice test on the twenty selected target words of the text were administered to all children. The participants were also submitted to a vocabulary (WASI) and reading test (Word/Pseudoword Reading Task) to evaluate their skills. In these final tests, children score better in words that had been tested in comparison with words that had been restudied. The results suggest that the retrieval practice is highly beneficial for enhancing the memory retention of education materials in children.

Keywords: retrieval practice; testing effect; memory; learning; education

4.1 Introduction

The retrieval practice is the act of recalling or retrieving information from memory rather than restudying or rereading it (Bjork, 1988). Its success is related to the idea of **making an effort from within** (Roediger & Butler 2011) to produce better memory retention. The retrieval practice has been studied by an increasing number of researchers in the area of education and memory. Several articles (e.g., Carrier & Pashler, 1992; Carpenter & DeLosh, 2005; Roediger & Karpicke, 2006; Agarwal et al. 2017) have reported the effect of this practice, showing positive results in experiments conducted in laboratories or in classroom environments.

Regarding the efficacy of the retrieval practice in enhancing students' learning, numerous studies (Callender & McDaniel, 2009; McDaniel &Callender, 2008; Karpicke and Roediger III, 2009; Roediger & Butler, 2011) were done contrasting this practice with rereading and restudying. Karpicke and Roediger (2009), for instance, conducted a survey and asked the participants to list their studying strategies, and choose between rereading or self-testing after reading a certain book chapter. Just a few students opted for the retrieval practice, and the authors concluded that the participants who did not engage in self-testing strategies experienced what they called **illusions of competence**, that is, a false impression that the content reread had been assimilated. This means that it is necessary to make the students themselves aware of the effectiveness of retrieval practice as a means of improving their learning development. In addition, Roediger & Butler (2011) show, in three experiments with students from middle school, that, since testing requires effortful retrieval, it is through this practice that education performance is enhanced. To prove this, the researchers evaluated students' performance through items which were quizzed and not-quizzed during the semester in chapter and unit exams. The results suggest the retrieval practice can boost education performance in real education setting.

The effectiveness of the retrieval practice is related to the effort that is made at the time of recalling. Many researchers (Pashler, Harold et al., 2007; Karpicke & Roediger, 2008; Roediger & Butler, 2011) argued that the act of recovering information requires mental effort, and, through this practice, learning is consolidated and the information can be recovered in the future. Pyc & Rawson (2009) have conducted two experiments to verify **the retrieval effort hypothesis**, which claims that difficult retrievals are better for memory than easier ones, since the effort involved in the task is better for memory. They have come to a conclusion that confirms the hypothesis, that

is, difficult retrieval stimulate higher levels of memory.

Various studies have investigated the retrieval practice in children, mostly in laboratory settings (Bouwmester & Verkoeijen, 2011; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Roher, Taylor, & Sholar, 2010). These reports have all corroborated the benefits of the retrieval practice, but the tasks involved in the experiments did not resemble real educational material used in the school environment. Other studies have verified if the retrieval practice can be replicated in the classroom context with real learning material (Carpenter et al.,2009; Lipko-Speed, Dunlosky, and Rawson, 2014; Roediger et al. 2011) and have successfully shown that this strategy can improve learning in such contexts.

However, few studies have conducted experiments with elementary students (Lipko-Speed et al., 2014; Jaeger et al., 2014, Goosens et al. 2014a). Since elementary school is such a relevant time in students' educational development, more studies should examine the possibility of using retrieval practice to teach these students (Karpicke, Blunt, Smith, & Karpicke, 2014). For instance, it is in this period that children begin to develop strategies that will help them learn from what they are reading. Jaeger et al. (2014) conducted a test in which third-grade children read a text and then performed an exam on selected target contents or reread the same text twice. They also had their IQ and reading skills tested and, after seven days, a four-choice memory test was administered in all participants. After the results were analyzed, they suggested that cued recall tests elicit robust testing effects in young children, since the students who had taken the test had significant better scores than those who just restudied the text twice.

In this article, we report an experiment with sixth-grade children. They initially read a brief encyclopedic text twice. Then they performed a cued recall test on ten of the selected target contents of the text and reread the other ten. After seven days, they performed a fill-in-the-blanks and a multiple-choice test on all the twenty target contents: the tested and the restudied. The students' reading and vocabulary skills were also investigated as we expect to generate possible associations between them and the effects of the retrieval practice.

As individual cognitive differences are claimed to influence learning (Unsworth & Engle, 2007), we expected to find that reading and vocabulary skills would have an import role in the performance of the memory tests. Thus, individuals with high vocabulary and reading skills were expected to have a better performance in the memory tests than those with lower ones.

4.2 Methods

4.2.1 Participants

The participants were 63 children with ages ranging from 11 years to 14 years, an average age of 11,86 (SD = 0,17). All children were enrolled in the sixth grade of one Brazilian elementary school (private) located in southeastern Brazil. Two children who were initially enrolled in the study were excluded due to not showing up for the second session of the experiment. Twenty-four students had their WASI results eliminated due to technical problems during the application. Informed consent was obtained in accordance with the institutional review board of the Federal University of Minas Gerais (UFMG), and the children's parents filled in signed consent forms.

4.2.2 Materials and stimuli

The study material was a text containing 335 words about the Sun, which was adapted from the material used by Jaeger, Eisenkraemer & Stein (2014), and can be seen in attachment 1. Twenty 'target facts' were selected from the text and used as probes for the memory tasks. Each target fact consisted of a word within a sentence of the text. For example, the word 'solis' in the sentence 'The word sun derives from the Latin word solis', and the word 'system' in the sentence 'The sun is the central star of our solar system'. To assess whether the text was appropriate for sixth-grade students, their own Portuguese teacher assessed the appropriateness of the subject, language, vocabulary, formal structure, and length of the text for the participants.

Two tests were administered to assess the children's reading skills. The first was the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI). The second was the Word/Pseudoword Reading Task developed by Salles et al. (2013). In this task, children read aloud regular, irregular, and pseudowords (20 of each type), and total reading time and accuracy are registered. In the current study, however, instead of presenting all words and pseudowords at once, we presented them individually in the center of a computer screen (Times New Roman size 50 white ink over a black screen). Each word/ pseudowords remained on screen for 2000 milliseconds (ms), and was followed by a blank screen (500 ms) before the appearance of the nex

word/ pseudowords. Children were asked to read each word/ pseudowords aloud as fast as they could, while being careful to pronounce them correctly. Accuracy in this task consisted in the correct pronunciation of the words and an appropriate pronunciation of the pseudowords, and was assessed by the experimenter during the task and checked afterwards whenever needed. In reading pseudowords, all reading possibilities are accepted as correct according to Brazilian Portuguese rules (e.g.,nefochoza, nefoksoza or nefossoza, erekêla, erequéla or erecuela). Regarding response times, a microphone connected to the computer registered the onset of each word pronunciation (i.e., the beginning of each word utterance) to the millisecond level. This task was programmed in the version 3 of the Psychophysics Toolbox for MATLAB (Brainard, 1997).

4.2.3 Procedures

The memory task comprised two sessions separated by an interval of seven days. All stages of the memory task were performed in the classroom, during school hours. During the first session, participants received three individual booklets, one for each part of that session. On the first booklet children were instructed to read the text "The Sun" twice (once by themselves and once along with the experimenter) and then performed a brief buffer task involving solving basic mathematical operations. This distractor, which was administered just after all the participants read the test twice, consisted of performing calculation using the four basic mathematics operations (addition, subtraction, multiplication and division). The grade of effectiveness of these particular operations was assessed by the school coordinator, which considered them to be "appropriate". Following this, the participants performed a cued recall test with half of the target-words (retrieval practice condition) and reread the remaining half target-words, which were now bolded (restudy condition). Note that as in the version of the text "The sun" presented earlier (first booklet), all target-words were embedded within the text (see attachment 2). Once all children finished reading and completing the blanks, the test was over. No feedback was provided to the students during or after the cued recall test.

After an interval of seven days, all participants performed a fill-in-the-blanks test. The task consisted of completing the 20 phrases extracted from the text with the respective missing target-words (see attachment 3). Participants were instructed to try as hard as they could to remember the alternative that, according to the text "The Sun", completed the sentences correctly. After finishing this task, students received a multiple-choice recognition task with the

studied/tested target words. This task consisted of completing the 20 phrases extracted from the text with the respective target facts, which were presented as one among four alternatives (see attachment 4). The three-distracter alternatives were similar to the target words and equally plausible for those with no previous knowledge of the subject (e.g. 'The superficial layer of the sun is called _____: (a) troposphere, (b) lithosphere, (c)photosphere, (d) stratosphere'). Children had up to thirty minutes to complete this task. No feedback was provided to the students during or after any of the tests.

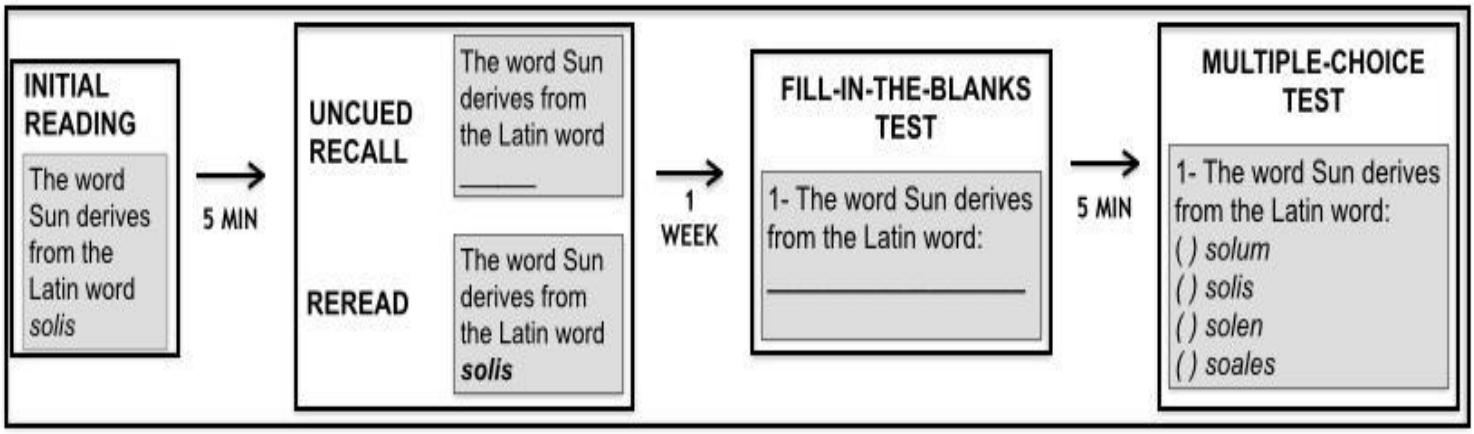


Figure1 -Illustration diagram of the experiment

4.3 Results

We first report the overall results for the WASI and the LPI, followed by the results for the retrieval practice administered in the first day of the experiment. Then we examined whether the typical testing effect was elicited, as well as the correlation analyzes between this effect and the individual vocabulary and reading scores. For all memory analyzes, one point was assigned to each correct response, and zero for each blank or incorrect response (maximum score in each memory test is 10). For the correlation analyzes, we implemented the testing effect as the subtraction of the restudy scores from the test scores. Thus, for each participant, we ended up with an index of the testing effect that was actually the advantage (or disadvantage) of test over restudy.

Children obtained an average score of 64.06 in the WASI vocabulary subtest, with results ranging from 38 to 80 (SD = 11.41). Considering the LPI, they obtained a mean accuracy of 56.8,

(94.66%) with scores ranging from 45 to 60 in this test (SD = 2.45). Their average mean reaction time (mRT) in this test was of 754.07 miliseconds (ms), with scores varying from 569.38 to 1155.6 ms (SD = 135.6 ms), which shows a great variability. In the initial memory test (i.e., retrieval practice), children recalled on average 7.44 (SD = 2.72) target facts. Considering that the maximum score was 10 target facts, this indicates that although children recalled most target facts, a certain level of difficulty was produced by the task, which is an important condition for the effectiveness of retrieval practice (Pyc & Rawson 2009).

Seven days later, children recalled on average 5.79 restudied words (SD = 2.19) and 6.61 words tested in the fill-in-the-blanks test (SD = 2.15). A t-test demonstrated that this advantage for test over restudy was statistically significant, **t**(60) = 2.55, $p < 0.02$, **d** = 0.66. This pattern was replicated for the multiple-choice test, wherein children recognized on average 8.30 restudied words (SD = 1.40) and 8.68 tested words (SD = 1.41), **t**(60) = 2.23, $p < 0.03$, **d** = 0.57. These results show that the current paradigm produced the expected "testing effect", and that this occurred for both types of tests. The question then is whether this effect is differently produced according to children's vocabulary and reading skills.

Correlations between the WASI's vocabulary subtest and the testing effect elicited by practicing retrieval with fill-in-the-blanks tests and multiple-choice tests were low (**r** = 0.012 and **r** = 0.113 respectively). Correlations between LPI's mRT and accuracy and the testing effect elicited by practicing retrieval with the multiple-choice test also low (r = 0.104 and r = 0.158 respectively), as well as the correlation between LPI's accuracy scores and the memory performance in the fill-in-the-blanks test (r = 0.102). Therefore, reading skills seem to not have a relationship with memory performance, and the retrieval practice can be effective independently of reading level of the participant.

4.4 Discussion

The present study investigated the efficiency of the testing effect in children and its association with reading and vocabulary skills. All participants in the current study were enrolled in the sixth grade. The material consisted of a text with twenty selected words that was read by all participants twice. After five minutes, they completed a cue-recall test on half of the selected target facts and reread the other half. Seven days later, the students took a multiple-choice test

and a sentence completion one in which a significant number of them showed a better performance for tested items than for restudied items.

Alongside these tests, the participants had their vocabulary and reading skills assessed. The first was the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI) more specifically the vocabulary subtest, which composes the Verbal Scale. As stated, the effectiveness of the testing effect is not related to a specific vocabulary performance, the retrieval practice occurs in participants with varied vocabulary skills. That is, in general, children showed positive effects of retrieval practice regardless of their performance on the vocabulary comprehension.

The second was the Word/Pseudoword Reading Task developed by Salles **et al** (2013). As already stated above, the correlation between reading skills and memory performance was not significant. This shows that the retrieval practice bears no relation to high or low reading skills, that is, the testing effect occurs independently of the participants' reading performance.

This result also corroborates other findings (Goossens et al. 2014, Karpicke et al. 2016) which suggest that children consistently benefit from retrieval practice. This effect is approximately the same regardless of reading comprehension or vocabulary level, which means that there is no evidence that retrieval practice is moderated by cognitive factors such as reading or vocabulary skills. For instance, Karpicke et al. 2016 reported that practicing retrieval may be an effective learning strategy for children with varying levels of reading comprehension and processing speed. In their study, the participants' individual differences in reading comprehension and processing speed were assessed and counterbalanced with their performance in three experiments. As a result, they observed that retrieval practice enhanced retention relative to repeated study with elementary school children and that this effect does not appear to depend on reading comprehension or processing speed.

The present findings are similar to those stated by Goossens et al. (2014a). They found benefits of retrieval practice with feedback in elementary school children who learned vocabulary words. Goossens et al. (2014a) reported that individual differences on a standardized test of children's vocabulary size test did not interact with the size of the retrieval practice effect. Just as the benefits of retrieval practice were similar for children with high or low vocabulary sizes, the present results suggest the benefits are similar regardless of whether children have high or low reading comprehension and processing speed scores.

In contrast to other studies investigating the testing effect in children (Carpenter et al., 2009; McDaniel et al., 2011; Roediger& Butler, 2011), feedback was not presented to the participants in the current experiment. Although providing feedback when participants are retrieving studied information increases the testing effect, (McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006), the results encountered here suggest that it is not vital to stimulate the retrieval practice (see also Jaeger et al. 2014), and tests without feedback (as the paper-and-pencil cued recall test administered here) might also be important tools to improve learning in educational contexts.

The power of retrieval practice in consolidating memory has been widely discussed and it has been proved that it is more efficient than just studying and rereading the material. As already verified in other articles (Carpenter et al., 2009; Karpicke & Roediger, 2007; Roediger & Karpicke, 2006; McDaniel et al., 2011; Roediger & Butler, 2011), this experiment provides strong evidence that validates the testing effect in a real education environment. The retrieval practice found in this study is in accord to the necessity to try to improve memory and learning. By trying to recall the information, the testing effect enhances memory more than reading or restudying some information. Cranney et al. (2009), Larsen et al. (2009) and McDermott at al. (2014) have demonstrated that implementing the retrieval practice through short quizzes is more effective than just reviewing the content. These studies involved a range of school environments, middle school, undergraduate psychology classes and medical school classrooms, and the results corroborates the statement that repeated retrieval practice is more efficient in improving long-term learning and comprehension than reading the information repeatedly. It is necessary to make students aware that studying regularly through frequent quizzing or self-testing, for instance, lead to better performance in final exams and increase retention of specific information.

We showed that retrieval practice is highly beneficial for enhancing the memory retention in children. In addition to findings from prior studies investigating this effect (Carpenter et al., 2009; McDaniel et al., 2011; Roediger & Butler, 2011), the present study reveals the efficiency of the testing effect for children who are enrolled in the sixth grade independently of their levels of vocabulary and reading skills. Further research is needed to directly assess the extent to which the benefits of retrieval practice depend on the characteristics of individual learners. It is possible that with a wider range of individual difference scores and additional children, there may be

individual difference factors that matter for retrieval practice effects.

4.5 References

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger III, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. **Memory**, **25**(6), 764-771.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. **Practical aspects of memory: Current research and issues**, **1**, 396-401.

Bouwmeester, S., & Verkoeijen, P. P. (2011). The effect of instruction method and relearning on Dutch spelling performance of third-through fifth-graders. European journal of psychology of education, 26(1), 61-74.

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. **Contemporary Educational Psychology**, **34**(1), 30-41.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. Applied Cognitive Psychology, 19(5), 619-636.

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., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. Applied Cognitive Psychology, 23, 760–771.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. Memory & Cognition, 20(6), 633-642.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. **European Journal of Cognitive Psychology**, **21**(6), 919-940.

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.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. **Paidéia (Ribeirão Preto)**, **23**(56), 397-406.

Fumagalli de Salles, J., da Rosa Piccolo, L., de Souza Zamo, R., & Toazza, R. (2013). Normas de desempenho em tarefa de leitura de palavras/pseudopalavras isoladas (LPI) para crianças de 1º ano a 7º ano. **Estudos e Pesquisas em Psicologia**, **13**(2).

Goossens, N. A., Camp, G., Verkoeijen, P. P., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. **Journal of Applied Research in Memory and Cognition**, **3**(3), 177-182.

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2014). Test-enhanced learning in third-grade children. **Educational Psychology**, **35**(4), 513-521.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. **science**, **319**(5865), 966-968.

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own?. Memory, 17(4), 471-479.

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. Current Directions in Psychological Science, 21, 157–163. doi:10.1177/ 0963721412443552

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. **Medical education**, **43**(12), 1174-1181.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science?. **Journal of Applied Research in Memory and Cognition**, **3**(3), 171-176.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. **Psychonomic bulletin & review**, **14**(2), 200-206.

McDaniel, M. A., & Callender, A. A. (2008). Cognition, memory, and education. In H. L. Roediger (Ed.), Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory (pp. 235–244). Amsterdam, The Netherlands: Elsevier.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. Journal of Educational Psychology, 103, 399–414. doi:10.1037/a0021782

McDermott, K. B., Agarwal, P. K., D'antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. **Journal of Experimental Psychology: Applied**, **20**(1), 3.

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.

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.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. **Journal of experimental psychology: Learning, Memory, and Cognition**, **21**(4), 803.

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

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. **Perspectives on Psychological Science**, **1**(3), 181-210.

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15, 20–27. doi:10.1016/j.tics.2010.09.003

Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. **Journal of Experimental Psychology: Applied**, **17**(4), 382

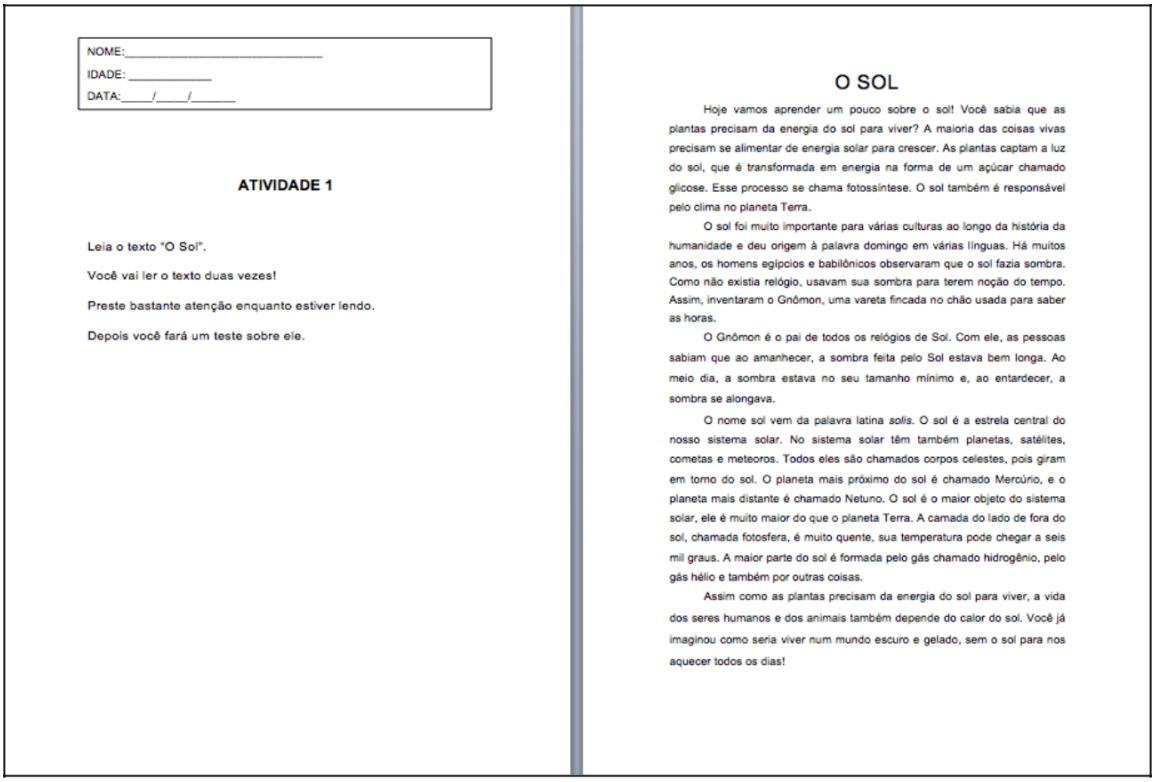
Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. **Journal of Applied Research in Memory and Cognition**, **1**(4), 242-248.

Roher, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. Journal of Experimental Psychology: Learning, Memory and Cognition, 36, 233–239. doi:10.1037/ a0017678

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. **Psychological review**, **114**(1), 104.

4.6 Attachments

Attachment 1: reading activity and distractor



NOME:_____
IDADE: _____
DATA:____/____/_____

ATIVIDADE 1

Leia o texto "O Sol".

Você vai ler o texto duas vezes!

Preste bastante atenção enquanto estiver lendo.

Depois você fará um teste sobre ele.

O SOL

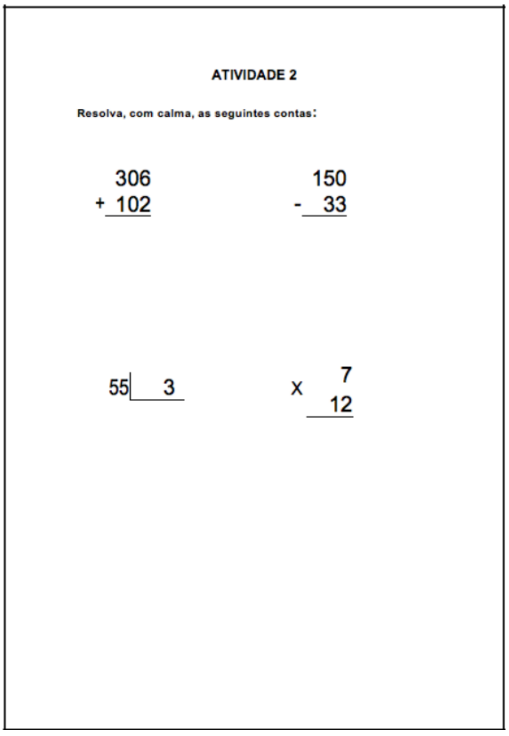
Hoje vamos aprender um pouco sobre o sol! Você sabia que as plantas precisam da energia do sol para viver? A maioria das coisas vivas precisam se alimentar de energia solar para crescer. As plantas captam a luz do sol, que é transformada em energia na forma de um açúcar chamado glicose. Esse processo se chama fotossíntese. O sol também é responsável pelo clima no planeta Terra.

O sol foi muito importante para várias culturas ao longo da história da humanidade e deu origem à palavra domingo em várias línguas. Há muitos anos, os homens egípcios e babilônicos observaram que o sol fazia sombra. Como não existia relógio, usavam sua sombra para terem noção do tempo. Assim, inventaram o Gnômon, uma vareta fincada no chão usada para saber as horas.

O Gnômon é o pai de todos os relógios de Sol. Com ele, as pessoas sabiam que ao amanhecer, a sombra feita pelo Sol estava bem longa. Ao meio dia, a sombra estava no seu tamanho mínimo e, ao entardecer, a sombra se alongava.

O nome sol vem da palavra latina *solis*. O sol é a estrela central do nosso sistema solar. No sistema solar têm também planetas, satélites, cometas e meteoros. Todos eles são chamados corpos celestes, pois giram em torno do sol. O planeta mais próximo do sol é chamado Mercúrio, e o planeta mais distante é chamado Netuno. O sol é o maior objeto do sistema solar, ele é muito maior do que o planeta Terra. A camada do lado de fora do sol, chamada fotosfera, é muito quente, sua temperatura pode chegar a seis mil graus. A maior parte do sol é formada pelo gás chamado hidrogênio, pelo gás hélio e também por outras coisas.

Assim como as plantas precisam da energia do sol para viver, a vida dos seres humanos e dos animais também depende do calor do sol. Você já imaginou como seria viver num mundo escuro e gelado, sem o sol para nos aquecer todos os dias!



ATIVIDADE 2

Resolva, com calma, as seguintes contas:

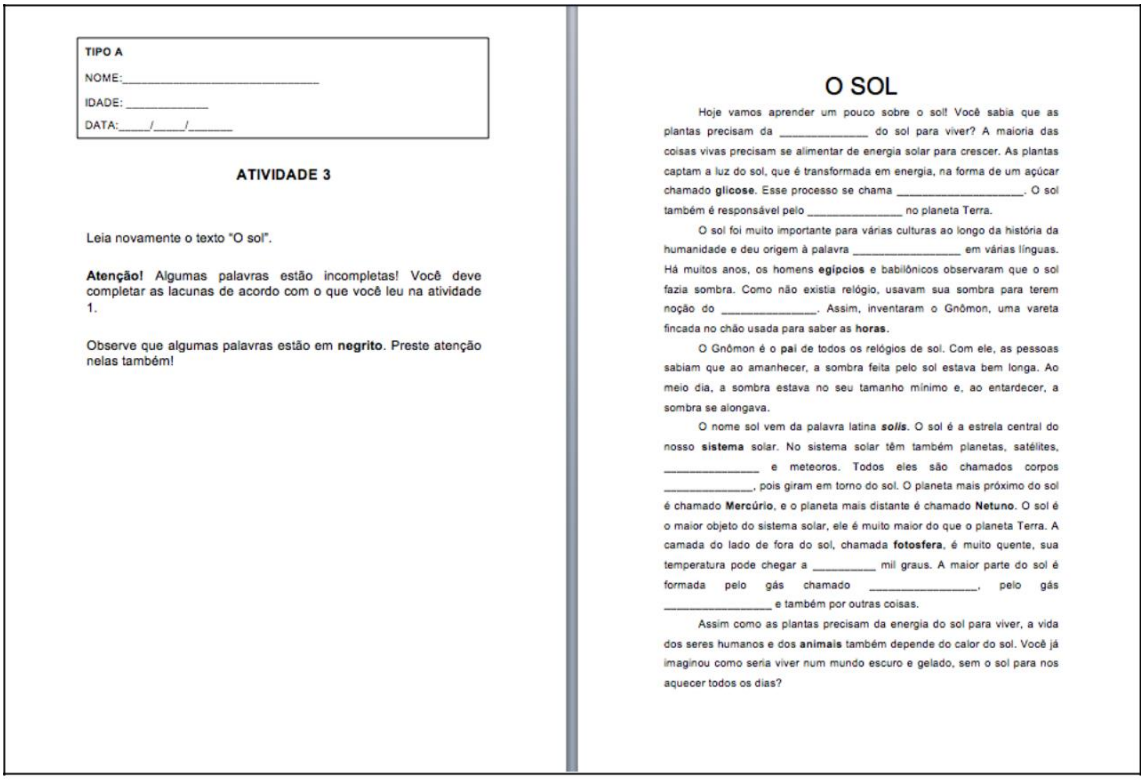
306 + 102

150 - 33

55 | 3

7 x 12

Attachment 2: restudying and testing activities (types A and B)



TIPO A

NOME:_____

IDADE: _____

DATA:_____/_____/_____

ATIVIDADE 3

Leia novamente o texto "O sol".

Atenção! Algumas palavras estão incompletas! Você deve completar as lacunas de acordo com o que você leu na atividade 1.

Observe que algumas palavras estão em **negrito**. Preste atenção nelas também!

O SOL

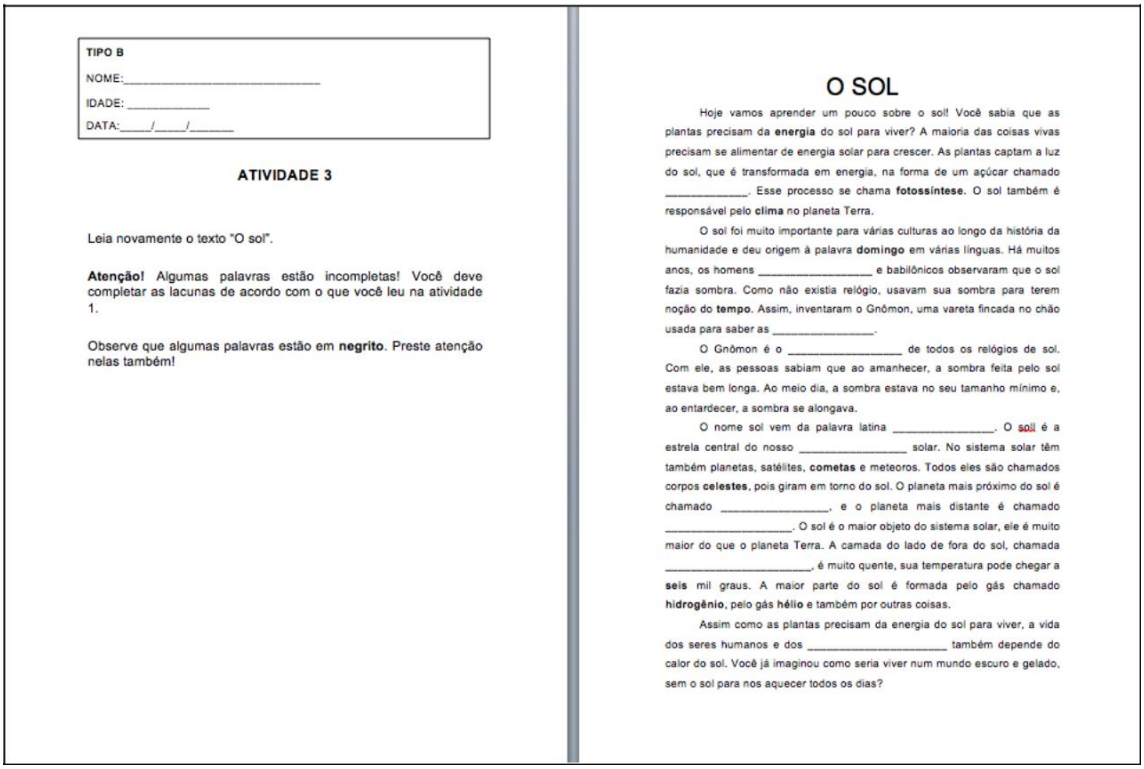
Hoje vamos aprender um pouco sobre o sol! Você sabia que as plantas precisam da _____ do sol para viver? A maioria das coisas vivas precisam se alimentar de energia solar para crescer. As plantas captam a luz do sol, que é transformada em energia, na forma de um açúcar chamado **glicose**. Esse processo se chama _____. O sol também é responsável pelo _____ no planeta Terra.

O sol foi muito importante para várias culturas ao longo da história da humanidade e deu origem à palavra _____ em várias línguas. Há muitos anos, os homens **egípcios** e babilônicos observaram que o sol fazia sombra. Como não existia relógio, usavam sua sombra para terem noção do _____. Assim, inventaram o Gnômon, uma vareta fincada no chão usada para saber as **horas**.

O Gnômon é o **pai** de todos os relógios de sol. Com ele, as pessoas sabiam que ao amanhecer, a sombra feita pelo sol estava bem longa. Ao meio dia, a sombra estava no seu tamanho mínimo e, ao entardecer, a sombra se alongava.

O nome sol vem da palavra latina ***solis***. O sol é a estrela central do nosso **sistema** solar. No sistema solar têm também planetas, satélites, _____ e meteoros. Todos eles são chamados corpos _____, pois giram em torno do sol. O planeta mais próximo do sol é chamado **Mercúrio**, e o planeta mais distante é chamado **Netuno**. O sol é o maior objeto do sistema solar, ele é muito maior do que o planeta Terra. A camada do lado de fora do sol, chamada **fotosfera**, é muito quente, sua temperatura pode chegar a _____ mil graus. A maior parte do sol é formada pelo gás chamado _____, pelo gás _____ e também por outras coisas.

Assim como as plantas precisam da energia do sol para viver, a vida dos seres humanos e dos **animais** também depende do calor do sol. Você já imaginou como seria viver num mundo escuro e gelado, sem o sol para nos aquecer todos os dias?



TIPO B

NOME:_____

IDADE: _____

DATA:_____/_____/_____

ATIVIDADE 3

Leia novamente o texto "O sol".

Atenção! Algumas palavras estão incompletas! Você deve completar as lacunas de acordo com o que você leu na atividade 1.

Observe que algumas palavras estão em **negrito**. Preste atenção nelas também!

O SOL

Hoje vamos aprender um pouco sobre o sol! Você sabia que as plantas precisam da **energia** do sol para viver? A maioria das coisas vivas precisam se alimentar de energia solar para crescer. As plantas captam a luz do sol, que é transformada em energia, na forma de um açúcar chamado _____. Esse processo se chama **fotossíntese**. O sol também é responsável pelo **clima** no planeta Terra.

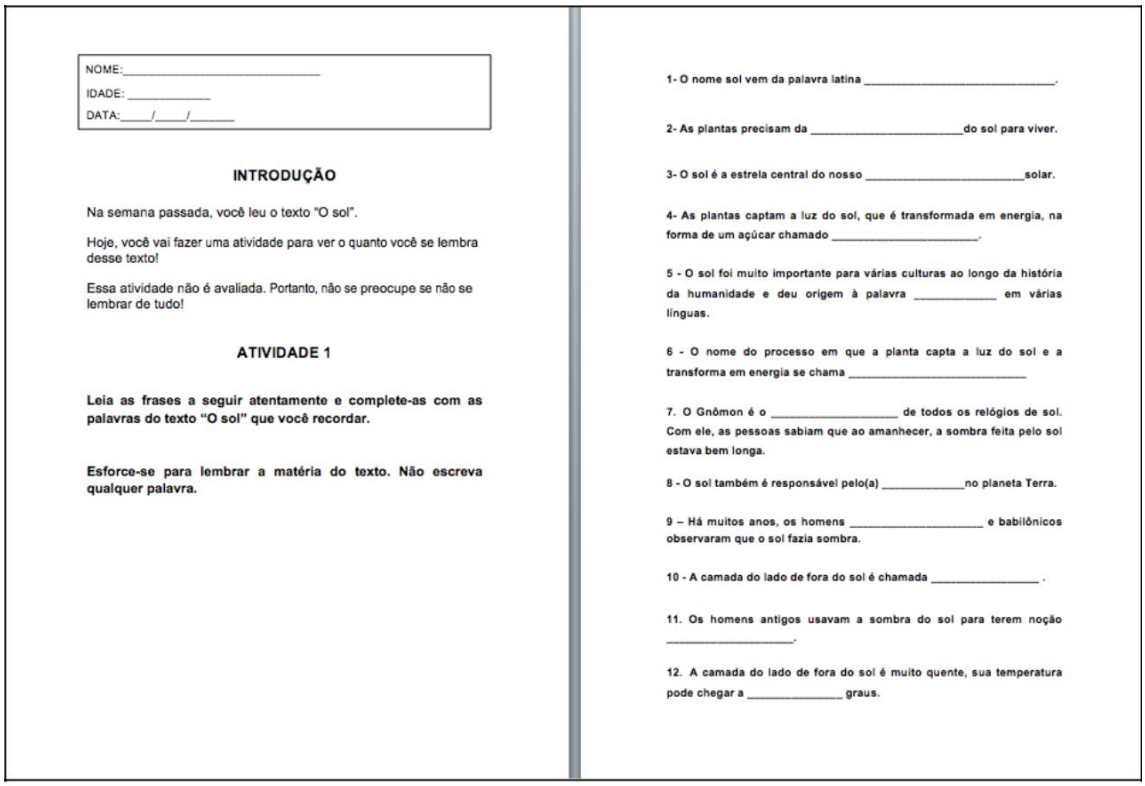
O sol foi muito importante para várias culturas ao longo da história da humanidade e deu origem à palavra **domingo** em várias línguas. Há muitos anos, os homens _____ e babilônicos observaram que o sol fazia sombra. Como não existia relógio, usavam sua sombra para terem noção do **tempo**. Assim, inventaram o Gnômon, uma vareta fincada no chão usada para saber as _____.

O Gnômon é o _____ de todos os relógios de sol. Com ele, as pessoas sabiam que ao amanhecer, a sombra feita pelo sol estava bem longa. Ao meio dia, a sombra estava no seu tamanho mínimo e, ao entardecer, a sombra se alongava.

O nome sol vem da palavra latina _____. O sol é a estrela central do nosso _____ solar. No sistema solar têm também planetas, satélites, **cometas** e meteoros. Todos eles são chamados corpos **celestes**, pois giram em torno do sol. O planeta mais próximo do sol é chamado _____, e o planeta mais distante é chamado _____. O sol é o maior objeto do sistema solar, ele é muito maior do que o planeta Terra. A camada do lado de fora do sol, chamada _____, é muito quente, sua temperatura pode chegar a **seis** mil graus. A maior parte do sol é formada pelo gás chamado **hidrogênio**, pelo gás **hélio** e também por outras coisas.

Assim como as plantas precisam da energia do sol para viver, a vida dos seres humanos e dos _____ também depende do calor do sol. Você já imaginou como seria viver num mundo escuro e gelado, sem o sol para nos aquecer todos os dias?

Attachment 3: fill-in-the-blanks test



NOME:_____
IDADE: _____
DATA:____/____/_____

INTRODUÇÃO

Na semana passada, você leu o texto "O sol".

Hoje, você vai fazer uma atividade para ver o quanto você se lembra desse texto!

Essa atividade não é avaliada. Portanto, não se preocupe se não se lembrar de tudo!

ATIVIDADE 1

Leia as frases a seguir atentamente e complete-as com as palavras do texto "O sol" que você recordar.

Esforce-se para lembrar a matéria do texto. Não escreva qualquer palavra.

1- O nome sol vem da palavra latina _____.

2- As plantas precisam da _____do sol para viver.

3- O sol é a estrela central do nosso _____solar.

4- As plantas captam a luz do sol, que é transformada em energia, na forma de um açúcar chamado _____.

5 - O sol foi muito importante para várias culturas ao longo da história da humanidade e deu origem à palavra _____ em várias línguas.

6 - O nome do processo em que a planta capta a luz do sol e a transforma em energia se chama _____

7. O Gnômon é o _____ de todos os relógios de sol. Com ele, as pessoas sabiam que ao amanhecer, a sombra feita pelo sol estava bem longa.

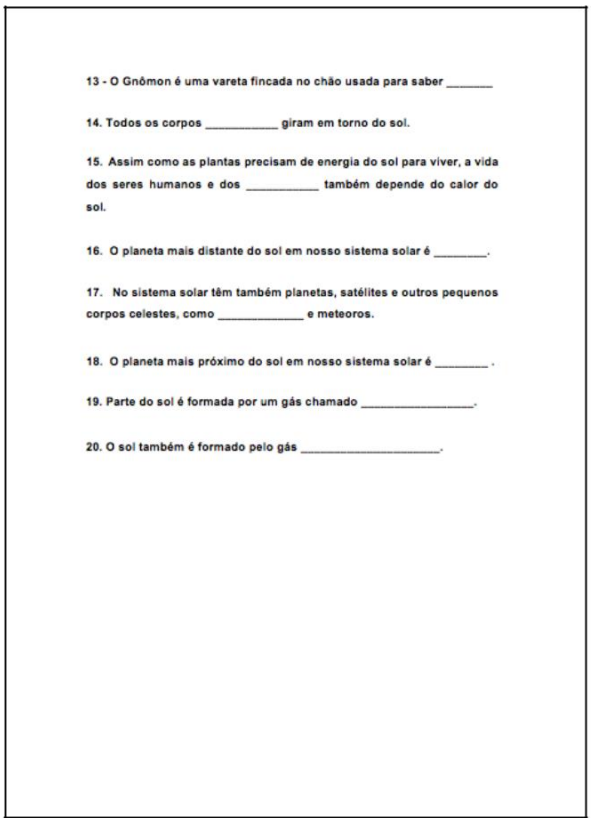
8 - O sol também é responsável pelo(a) _____no planeta Terra.

9 – Há muitos anos, os homens _____ e babilônicos observaram que o sol fazia sombra.

10 - A camada do lado de fora do sol é chamada _____ .

11. Os homens antigos usavam a sombra do sol para terem noção _____.

12. A camada do lado de fora do sol é muito quente, sua temperatura pode chegar a _____ graus.



13 - O Gnômon é uma vareta fincada no chão usada para saber _____

14. Todos os corpos _____ giram em torno do sol.

15. Assim como as plantas precisam de energia do sol para viver, a vida dos seres humanos e dos _____ também depende do calor do sol.

16. O planeta mais distante do sol em nosso sistema solar é _____.

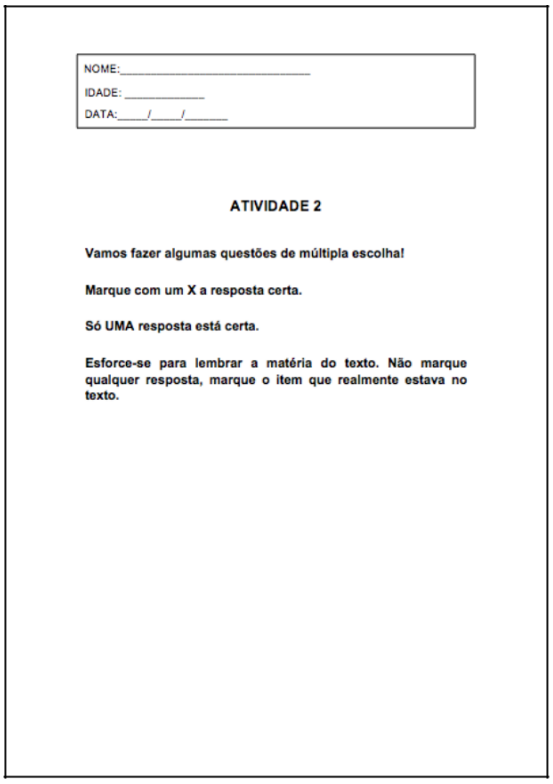
17. No sistema solar têm também planetas, satélites e outros pequenos corpos celestes, como _____ e meteoros.

18. O planeta mais próximo do sol em nosso sistema solar é _____ .

19. Parte do sol é formada por um gás chamado _____.

20. O sol também é formado pelo gás _____.

Attachment 4: Multiple-choice test



NOME:_____

IDADE: _____

DATA:____/____/_____

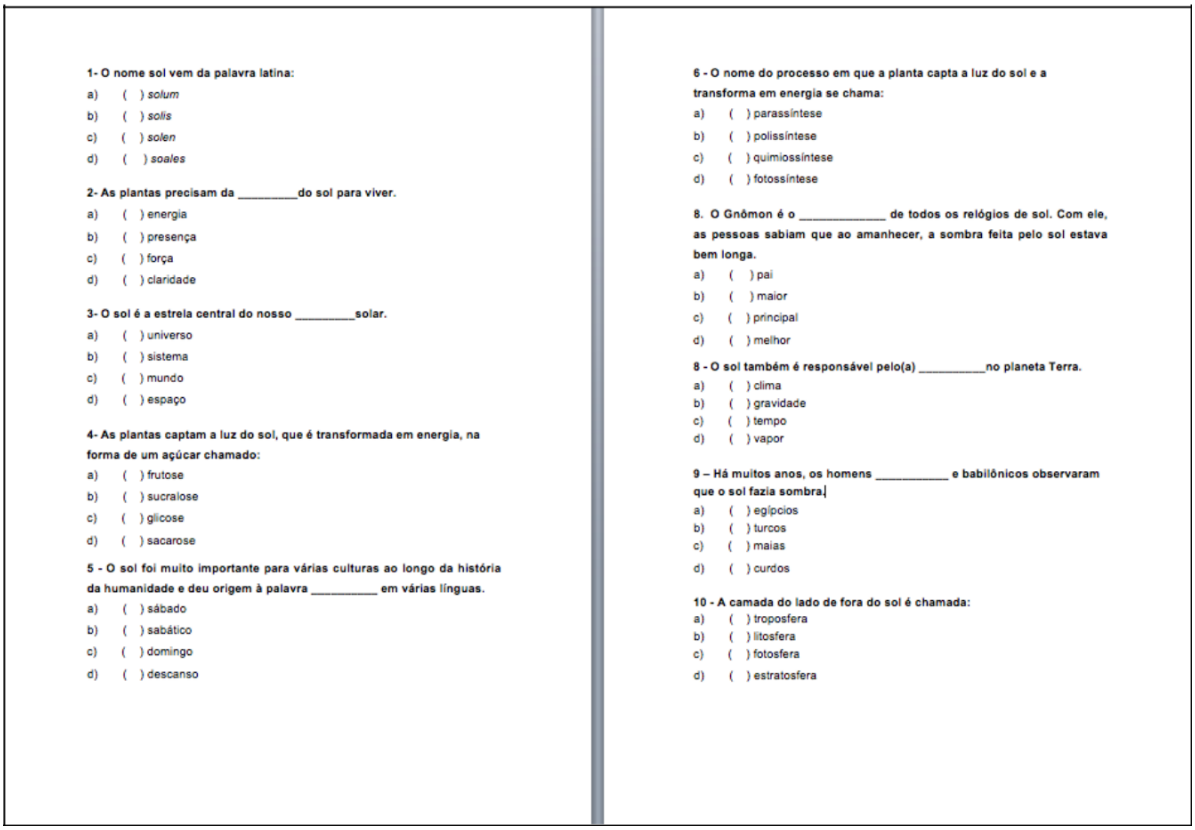
ATIVIDADE 2

Vamos fazer algumas questões de múltipla escolha!

Marque com um X a resposta certa.

Só UMA resposta está certa.

Esforce-se para lembrar a matéria do texto. Não marque qualquer resposta, marque o item que realmente estava no texto.



1- O nome sol vem da palavra latina:

a) ( ) *solum*
b) ( ) *solis*
c) ( ) *solen*
d) ( ) *soales*

2- As plantas precisam da _____do sol para viver.

a) ( ) energia
b) ( ) presença
c) ( ) força
d) ( ) claridade

3- O sol é a estrela central do nosso _____solar.

a) ( ) universo
b) ( ) sistema
c) ( ) mundo
d) ( ) espaço

4- As plantas captam a luz do sol, que é transformada em energia, na forma de um açúcar chamado:

a) ( ) frutose
b) ( ) sucralose
c) ( ) glicose
d) ( ) sacarose

5 - O sol foi muito importante para várias culturas ao longo da história da humanidade e deu origem à palavra _____ em várias línguas.

a) ( ) sábado
b) ( ) sabático
c) ( ) domingo
d) ( ) descanso

6 - O nome do processo em que a planta capta a luz do sol e a transforma em energia se chama:

a) ( ) parassíntese
b) ( ) polissíntese
c) ( ) quimiossíntese
d) ( ) fotossíntese

8. O Gnômon é o _____ de todos os relógios de sol. Com ele, as pessoas sabiam que ao amanhecer, a sombra feita pelo sol estava bem longa.

a) ( ) pai
b) ( ) maior
c) ( ) principal
d) ( ) melhor

8 - O sol também é responsável pelo(a) _____no planeta Terra.

a) ( ) clima
b) ( ) gravidade
c) ( ) tempo
d) ( ) vapor

9 – Há muitos anos, os homens _____ e babilônicos observaram que o sol fazia sombra.

a) ( ) egípcios
b) ( ) turcos
c) ( ) maias
d) ( ) curdos

10 - A camada do lado de fora do sol é chamada:

a) ( ) troposfera
b) ( ) litosfera
c) ( ) fotosfera
d) ( ) estratosfera

11. Os homens antigos usavam a sombra do sol para terem noção

a) ( ) da temperatura
b) ( ) da localização
c) ( ) da atmosfera
d) ( ) do tempo

13. A camada do lado de fora do sol é muito quente, sua temperatura pode chegar a _____ graus.

a) ( ) sete mil
b) ( ) seis mil
c) ( ) oito mil
d) ( ) cinco mil

13 - O Gnômon é uma vareta fincada no chão usada para saber:

a) ( ) os anos
b) ( ) as horas
c) ( ) os segundos
d) ( ) os meses

14. Todos os corpos _____ giram em torno do sol.

a) ( ) cósmicos
b) ( ) solares
c) ( ) lunares
d) ( ) celestes

16. Assim como as plantas precisam de energia do sol para viver, a vida dos seres humanos e dos _____ também depende do calor do sol.

a) ( ) mamíferos
b) ( ) lagartos
c) ( ) moluscos
d) ( ) animais

17. Qual o planeta mais distante do sol em nosso sistema solar?

a) ( ) Júpiter
b) ( ) Netuno
c) ( ) Terra
d) ( ) Marte

18. No sistema solar têm também planetas, satélites e outros pequenos corpos celestes, como _____ e meteoros.

a) ( ) asteroides
b) ( ) pulsares
c) ( ) cometas
d) ( ) estrelas

19. Qual o planeta mais próximo do sol em nosso sistema solar?

a) ( ) Plutão
b) ( ) Saturno
c) ( ) Mercúrio
d) ( ) Vênus

19. Parte do sol é formada por um gás chamado:

a) ( ) oxigênio
b) ( ) hidrogênio
c) ( ) argônio
d) ( ) neônio

20. O sol também é formado pelo gás:

a) ( ) lítio
b) ( ) nitrogênio
c) ( ) hélio
d) ( ) berílio

5. Conclusion

The current thesis analyzed the retrieval practice hypothesis in education settings. The first study reviewed a series of articles that examined the use of retrieval practice as a learning strategy in authentic educational contexts. The reviewed studies indicated that the retrieval practice consolidates memory and increase long-term retention, especially when compared to repeated studying**.** The benefits of this practice were noticed independently of time intervals and types of materials used in the tasks, as already discussed. As to feedback, it was not seen essential to promote the retrieval practice, although its presence is considered extremely valuable.

All data studied corroborated the fact that simply rereading or restudying some information produces little or no memory benefit while retrieval activities promote superior long-term retention. The experiments analyzed involved middle and high school students as well as undergraduate ones, and they all pointed out an improvement in the learner's knowledge proceeded from repeated quizzing and delayed tests**.** Thus, it is incontestable that the retrieval practice is not just a laboratory phenomenon.

The second study showed that the retrieval practice is highly beneficial for promoting long-term retention and learning of class relevant material in children. (Carpenter et al., 2009; McDaniel et al., 2011; Roediger et al., 2011; Jaeger et al., 2015). The present study also investigated potential associations between reading and vocabulary skills with the testing effect and did not find any correlation between these skills and performance in the later fill-in-the-blanks and multiple-choice tests.

In this experiment, no feedback was provided concerning the answers of the tests performed by the students. Although other studies have shown that feedback enhances the benefits of the retrieval practice (McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006), it is not considered fundamental to produce strong testing effects. However, it is important to emphasize that it is a relevant learning tool, since it not only promotes later retention, but also allows learners to correct metacognitive errors (Butler, Karpicke, & Roediger 2008).

Therefore, the retrieval practice hypothesis discussed in the review was also corroborated by the findings of the experiment in the second study. The repeated retrieval technique consolidates learning and promotes long-term retention. To propagate this idea into the educational scenario, it is essential that teachers and students become aware of its benefits.

Indeed, quizzes are seldom used as tools to prepare to final tests since their mnemonic effects are hardly ever known by faculty and students (Karpicke & Roediger, 2008). Therefore, it is important to show that repeated retrieval activities, when used correctly, help not only to access knowledge but also to consolidate it, making it possible to transfer knowledge to new situations. Survey questions given to students who have taken part in experiments evaluating the benefits of the retrieval practice often show that they feel more confident and prepared to take the final tests after being frequently quizzed (Roediger & Butler, 2011). Taking this into account, it is essential that teachers start substituting simple rereading and restudying activities for ones that promote repeated retrieval, as tests and quizzes.

In conclusion, the present thesis provides evidence suggesting the efficiency of the retrieval practice in the educational setting. In addition, it presents a substantial review of studies with learners of several school ages, discussing various aspects such as feedback, type of material and their implication to the success of the retrieval practice technique.

5.1 References

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. **Applied Cognitive Psychology, 23,** 760-771.

Carpenter, S. K., Lund, T. J., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. **Educational Psychology Review**, **28**(2), 353-375.

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,** 4-58.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. **Paideia, 23,** 397-406.

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2015). Test-enhanced learning in third-grade children. **Educational Psychology, 35,** 513-521.

Karpicke, J. D. (2012). Retrieval-based learning: active retrieval promotes meaningful learning. **Current Directions in Psychological Science, 21,** 157–163.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. **Science, 319,** 966-968.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. **Psychonomic bulletin & review**, **14**(2), 200-206.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., Roediger, H. L. (2011). Test-enhanced learning in a middle school classroom: the effects of quiz frequency and placement. **Journal of Educational Psychology, 103,** 399–414.

McDermott, K. B., Agarwal, P. K., D'antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. **Journal of Experimental Psychology: Applied**, **20**(1), 3.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. **Perspectives on Psychological Science, 1,** 181–210.

Roediger, H. L., Agarwal, P. K.; McDaniel, M. A.; McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. **Journal of Experimental Psychology: Applied, 17,** 382-395.

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

Roher, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. Journal of Experimental Psychology: Learning, Memory and Cognition, 36, 233–239. doi:10.1037/ a0017678