UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM NEUROCIÊNCIAS

THUAN HENRIQUE PEDROSA GOMES MARTINS

IS RETRIEVAL PRACTICE BETTER THAN RESTUDY WITH EFFICIENT ENCODING STRATEGIES?

BELO HORIZONTE
2022

THUAN HENRIQUE PEDROSA GOMES MARTINS

IS RETRIEVAL PRACTICE BETTER THAN RESTUDY WITH EFFICIENT ENCODING STRATEGIES?

Versão Final

Dissertação apresentada ao Programa de Pós-Graduação *Strictu Sensu* em *Neurociências* do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Neurociências

Orientador: Prof. Dr. Antônio Jaeger

BELO HORIZONTE
2022

043 Martins, Thuan Henrique Pedrosa Gomes.
Is retrieval practice better than restudy with efficient encoding strategies? [manuscrito] / Thuan Henrique Pedrosa Gomes Martins. – 2022.
45 f.: il. ; 29,5 cm.

Orientador: Prof. Dr. Antônio Jaeger.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-graduação em Neurociências.

1. Neurociências. 2. Testes de Memória e Aprendizagem. I. Jaeger, Antônio. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 612.8

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis. CRB 6 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM NEUROCIÊNCIAS

FOLHA DE APROVAÇÃO

Is retrieval practice better than restudy with efficient encoding strategies?

THUAN HENRIQUE PEDROSA GOMES MARTINS

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em NEUROCIÊNCIAS, como requisito para obtenção do grau de Mestre em NEUROCIÊNCIAS, área de concentração NEUROCIÊNCIAS BÁSICAS.

Aprovada em 24 de março de 2022, pela banca constituída pelos membros:

Prof(a). Julia Beatriz Lopes Silva
UFMG

Prof(a). Renato Bortoloti
UFMG

Prof(a). Antonio Jaeger - Orientador
UFMG

Belo Horizonte, 24 de março de 2022.

Documento assinado eletronicamente por **Julia Beatriz Lopes Silva, Subcoordenador(a)**, em 24/03/2022, às 13:34, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Antonio Jaeger, Professor do Magistério Superior**, em 24/03/2022, às 14:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Renato Bortoloti, Professor do Magistério Superior**, em 24/03/2022, às 15:17, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1306899** e o código CRC **7F69690B**.

Referência: Processo nº 23072.213615/2022-70 SEI nº 1306899

AGRADECIMENTOS

Gostaria de agradecer aos meus pais pela vida e pela formação e dedicação ao longo de todos esses anos. À minha mãe pela perseverança e pelo raciocínio quente e ao meu pai pela calma e pelo raciocínio frio. Ao Ítalo agradeço pelas discussões, reflexões, críticas e ensinamentos, também pelo caminho trilhado e por ter dado os tão difíceis primeiros passos - me apresentando ao mesmo tempo a vida acadêmica. Ao Mário Lúcio também a companhia intelectual, que vem acompanhada da lealdade, do amor e da força que lhe são características e que são fundamentais à nossa família. Ganhei mais dois irmãos nos últimos anos pelos quais sou também grato.

À minha companheira Maria Luiza, que esteve comigo por quase toda essa trajetória agradeço ao apoio em todos os âmbitos e à família pela acolhida. Foi com ela que passei os momentos mais duros e mais prazerosos durante o meu mestrado, com ela também, ombro a ombro, que venho fazendo talvez a parte mais marcante da minha vida adulta.

Devo também aos meus mestres e professores Antônio Jaeger e Vitor Haase a minha iniciação e formação na Psicologia, área que aprendo cada dia mais a amar, tenho uma profunda dívida intelectual com ambos. A tenho também com o Professor Bernardo Vianna, que foi meu pilar na formação clínica, acadêmica e moral.

O companheirismo e as inspirações vêm constantemente das discussões que desenvolvi com os colegas da Psicologia Cognitiva e Neurociência Cognitiva. Uma das figuras a quem devo nesse sentido é o meu companheiro João Pedro Parreira. Entusiasta, como eu, dos temas mais complexos da Psicologia e que me ajudou a ter energia e esperança, mesmo em um cenário acadêmico tão desfavorável.

Agradeço também aos meus pacientes e colegas de trabalho pelos ensinamentos, que fazem com que a Psicologia esteja viva e presente a cada dia em minha vida e se renove**,** para se tornar cada vez mais algo próximo e claro para mim.

Resumo

O efeito de teste ou da prática de evocação é o ganho no desempenho obtido quando utilizamos como estratégia de aprendizagem a prática de evocação, quando comparada a outras estratégias, como o reestudo. Poucos trabalhos exploraram o efeito ou sua ausência quando se compararam a prática de evocação com outras estratégias. Nos experimentos aqui apresentados, nós comparamos dois níveis de processamento e por consequência, duas estratégias de aprendizagem com a prática de evocação. Nossa amostra foi composta de estudantes universitários ativos. O paradigma experimental foi instrasujeito, em que os sujeitos usaram três estratégias diferentes para aprenderem pares de palavras. Além da comparação entre as estratégias de aprendizagem, comparamos o desempenho em uma tarefa de memória associativa. Usamos o desempenho final em uma tarefa de reconhecimento e em uma tarefa de julgamento de associação para fazer as comparações. No primeiro experimento, os sujeitos fizeram um julgamento de consoantes, imagens mentais e a prática de evocação. No segundo experimento nós tentamos replicar os resultados obtidos no primeiro experimento usando julgamento de consoantes, pensamento autorreferenciado e a prática de evocação. Não foi encontrado efeito de teste ou de prática de evocação entre o reestudo com processamento profundo e a prática de evocação, as duas estratégias obtiveram uma performance superior quando comparadas ao reestudo com processamento superficial – julgamento de consoantes. Importante ressaltar que o processamento profundo obteve um desempenho superior em medidas que envolvem a memória para associação de informações, com os sujeitos lembrando uma quantidade maior de associações dos pares no Experimento 1. Esses resultados contrastam com os resultados comumente encontrados em outros trabalhos e reforçam a importância de encontramos estratégias boas e fortes para utilizarmos como condição controle quando se exploram os possíveis benefícios da prática de evocação, assim como os processos que subjazem o conhecido Efeito de Teste.

Palavras-chave memória, aprendizagem, efeito de teste, processamento profundo

Abstract

The testing effect is the enhancement of memory retention typically produced by the comparison of retrieval practice to other encoding strategies, such as restudy. Few studies compared retrieval practice with learning strategies known to produce stronger memory retention than restudy. Here, we compared two different levels of processing and by consequence, different learning strategies with retrieval practice. Our sample was composed of currently active college students. The paradigm was a within-subject in which subjects used three different strategies for learning word pairs. In addition to the comparison of the learning strategies, we compared the performance in an associative memory task. We used the final performance in a recognition task and in a judgment of association task to make the comparisons. In the first experiment, the subjects made a consonants judgment, a visual imagery task, and retrieval practice. In the second experiment, we tried to replicate the results obtained in the first experiment using consonants judgment, a self-referenced thought strategy, and retrieval practice. No testing or retrieval practice effect was found between retrieval practice and restudy with deep processing, both conditions obtained a better performance when compared to restudy with shallow processing – consonants judgment. Importantly, when it comes to associative memory, deep processing exhibited a better result, with subjects remembering more word pairs associations in Experiment 1. These results contrast the common results found in other studies and reinforce the importance to find good or strong control conditions when exploring the benefits of retrieval practice and the processes that underlie the well-known Testing Effect.

Keywords: memory, learning, testing effect, deep processing

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1. Introduction

The testing effect is the enhancement of memory retention typically obtained when retrieval practice is compared to other learning strategies, such as restudy. Although this phenomenon only became a major topic of research after the publication of a relatively recent study (Roediger & Karpicke, 2006a), it could be inferred from the first demonstrations of Hermann Ebbinghaus showing that it took fewer repetitions to learn a list of nonsense syllables after testing the same list (Ebbinghaus, 1885/1913). Retrieval practice might have benefited retention, a phenomenon that was observed in further historical periods of memory research (e.g., Spitzer, 1939). Currently, studies investigating the memory benefits of practicing retrieval typically adopt a design wherein a series of to-be-learned materials are initially studied, and after such initial study, the materials are either restudied (e.g., reread) or are subjected to tests requiring their retrieval (e.g., cued recall). After a retention interval that can vary from minutes to months, the learning materials are tested again, and the most common finding is greater memory retention for the materials that were retrieved than for the materials that were restudied.

The study which first championed this approach in the context of current memory research was published relatively recently (Roediger & Karpicke, 2006a), and helped the investigations on retrieval practice to flourish into experimental cognitive psychology. Using a control group, Roediger & Karpicke (2006a) showed that retrieval practice could generate memory benefits in a final test when compared with restudy. Since the study of Roediger & Karpicke (2006a) the testing effect was replicated in different contexts, with different materials and final test formats (Eisenkraemer, Jaeger & Stein, 2013; Karpicke & Roediger, 2006b; Rowland, 2014). The success of this learning strategy generated lots of enthusiasm and a large number of materials that aimed to influence, promote, and disseminate the retrieval practice in the educational context (Biwer et al, 2020; Ariel & Karpicke, 2018; McDermott,2021).

Although variations in the typical retrieval practice experimental paradigm are made with, for example, expanding intervals (Neuschatz et al.,2005; Morris et al., 2005) or multiple tests and restudy conditions (Roediger & Karpicke, 2006a), the main comparison of interest is the difference in the final performance between retrieval practice and restudy – that involves only the re-presentation of the information. Hence, the subjects are oriented to use the retrieval practice,

but there is not any specific strategy that is given for the restudy condition (e.g., Roediger & Karpicke, 2006a; Butler, 2010; Carpenter & Pashler, 2007).

This common comparison of the retrieval practice with restudy, however, presents many problems. The first problem is that restudy is not a strong and consistently used learning strategy, at least when we think of the classical paradigm that consisted of only showing the same material to the subject, that would only reread or reexperience the information. Restudy also does not allow the experimenter to control the strategy that the subject is using, when the subjects are restudying the items to-be-learned, they can use the strategy they want with the material. As shown in Minear et al. (2018) subjects tend to use many different strategies while experiencing the items.

The third problem is that the underpinning processes related to the testing effect remain a difficult problem to solve and finding further strategies that are as good as retrieval practice can not only help to solve this problem but could also contribute to the elaboration of good educational practices. Comparing retrieval practice with other strong strategies may reveal shared or dissociated mechanisms between retrieval practice and will allow testing the hypotheses that explain the retrieval practice effect. The fourth problem is that there is a controversy related to this effect, because some subjects may not beneficiate from this strategy as expected. In a current study reported by Minear et al. (2018), for example, one-third of the participants had a negative testing effect, namely, these subjects showed greater memory retention for restudied than for retrieved materials. Interestingly, these participants also reported using less “shallower” encoding learning strategies during restudy, a finding that is relevant for the purposes of the current study.

Despite the enthusiasm produced by the possibility of using retrieval practice as a learning strategy, few studies compared the retrieval practice with learning strategies other than restudy (Karpicke & Blunt, 2011; McDaniel et al., 2009; Morris et al., 2005; Neuschatz et al, 2005; Rummel et al., 2017; Stenlund et al., 2017). Hence, a possible comparison that emerges is comparing retrieval practice with “deep” and “shallow” processing learning strategies that, although hard to define, offer a great avenue for proposing new and strong strategies. The notion of depth of processing as having a role in memory retention was first proposed by Craik and Lockhart (1972), who defined deep processing as the semantic-based – or cognitive analytic-processing, and shallow processing as more superficial or attribute-based (i.e., with minimal involvement of semantic processing). Beyond the possibility of obtaining a good control

condition, the comparison of retrieval practice with deep processing strategies can help elucidate some of the underlying mechanisms of the testing effect, because many explanatory hypotheses can be tested with this comparison.

Su, Buchin & Mulligan (2020) investigated the importance of the levels of retrieval in the testing effect but did not directly explore the effects of levels of processing in restudying information – when compared with testing. A similar use of different levels of processing was made in the retrieval practice condition in Collins, Rasco & Benassi (2018). They compared the use of tests with varying levels of processing strategies in the second phase, mixing them with retrieval practice. They showed that tests with "deeper" processing strategies are better than tests with lower processing levels.

Other studies found inconsistent effects of testing, with retrieval practice not being as good as expected in some situations. In the study of McDaniel et al. (2009) taking notes had a similar result at the final test in multiple-choice, problem-solving, and short answer questions when compared to retrieval practice. Although Lechuga et al. (2015) found a testing effect, it was reduced when they compared the results of retrieval practice with students that were already familiarized with the use of conceptual maps. This is important because one of the possible advantages of retrieval practice is that it is easier to apply than techniques that are commonly compared – e.g., keyword-mnemonic and conceptual mapping. Other studies, such as Fritz et al (2007), found a different result: keyword mnemonic or its combination with retrieval practice were at least as good as retrieval practice. Neuschatz et al. (2005) found a marginally significant benefit over retrieval practice using name-face imagery and expanding rehearsal in their last experiment.

Other pieces of evidence are related to the combination of strategies. Combining conceptual maps and retrieval practice, in the study of O'Day & Karpicke (2020), did not promote any advantage over solely retrieving information. On the other hand, in the study of Neuschatz et al. (2005), the combination of name-face imagery and retrieval practice obtained a performance better than only retrieving information. In the study of Fritz et al. (2007), combining strategies neither promoted any benefit nor had a worse result – there was a marginally non-significant greater performance for combined strategies. Although the main goal of Karpicke & Smith (2012) was not to test the combination of strategies, they combined retrieval practice with keyword mnemonic and got a worse performance when the result in the combined condition was

compared with retrieval practice only. The same conclusion can be obtained in the study of Stenlund et al. (2017), the authors compared group discussions with retrieval practice in three tests (immediate, 1 week, and 4 weeks), this comparison led to a combination of strategies that were insufficient to recover the advantage obtained by retrieval practice in the study phase – and first detected in the immediate test.

Some studies used learning techniques that can be classified as deep processing strategies because they involve semantic-based – or cognitive analytic- processing, but they do not usually directly compare retrieval practice with these strategies. Chang (2017) explicitly tested the influence of the type of processing – shallow and deep- in the study phase over final performance, but after the study phase, all subjects did a retrieval practice in the second phase. Hence, the author did not compare strategies with different levels of processing with retrieval practice. Karpicke & Smith (2012) compared the keyword mnemonic and a visual imagery task with retrieval practice, but before using these learning strategies, subjects passed by a retrieval practice. In the former, there was a significantly better final performance for deep processing over shallow processing strategies. Karpicke & Smith (2012) found a retrieval practice effect.

Another technique that can be put in the same group is the name-face imagery. It is implemented using the face or a photo of a person, a prominent physical characteristic of the face, the name of the person, and a word phonologically similar to the person's name. The task is imagining a scene with the prominent feature and the word similar to the name and the main aim is learning the person's name (Neuschatz et al., 2005). Using only such name-face imagery, Neuschatz et al. (2005) found worse results in the final test when compared to retrieval practice for learning the association between names and faces. Even though the subjects must process and create images using the name, some facial characteristics, and words phonologically similar to the name, the association to-be-learned is arbitrary and one can question if it is a deep processing strategy, because the association is based on some similarities and involves many steps and not exactly the to-be-learned association.

A second technique that was compared with retrieval practice is the keyword mnemonic. It involves creating an image linking one word of interest to another using a keyword. For foreign languages, the subject will receive a pair of two correspondent words, one in their language and one with the same or similar meaning from another. They also will have access to a word in their own language that is phonologically similar to the foreign word (i.e., keyword). Then, they will

imagine a scene containing both the definition of the word to-be-learned and the keyword. The idea is to learn the foreign word and its meaning.

Fritz et al (2007) compared retrieval practice to keyword mnemonic for the learning of german words and found no advantage for retrieval practice. The same critique made above can be done in this case, however. That is, although this task is thought to elicit deep processing, the main relationship of the to-be-learned information and the equivalent in English is an arbitrary one. Both relationships, between the face and the name, in the study of Neuschatz et al. (2005) and Morris et al. (2005), and between the German word and the English word in the study of Fritz et al. (2007), are established by convention – put forth by the experimenter or based in language correspondence. The name of a person is not defined by its physical attributes or personality. Although there is a correspondence or similarity in the meaning of words in German and English, the equivalence can only be postulated by a person that speaks both languages and knows the meaning of both words, even then, their relationship cannot be inferred by simply studying them – unless they are cognates, in this case, it can be inferred by phonological or orthographical similarity.

Morris et al. (2005) used a semantic strategy that consisted of finding meaning on the name of a person, in which the photo was also present. They also used the name-face imagery technique in one of the experiments, but the results showed a testing effect when both techniques were compared to retrieval practice in two experiments. On the other hand, in their first experiment, there was no difference between retrieval practice and semantic association when they were used independently.

When it comes to conceptual mapping, using it as a restudy strategy can have many problems, it is a complex strategy because it depends on the expertise of the subjects (Lechuga et al, 2015), it also can facilitate the processing and maintenance of the information – because the subject can copy, write, and draw the information, reducing the amount of processing effort. Another problem is that it may take a long time to be applied (O'Day & Karpicke, 2020). Some of these problems explain why the testing effect can be observed when retrieval practice is compared to Conceptual Mapping (Karpicke & Blunt, 2011; Lechuga et al., 2015; O'Day & Karpicke, 2020).

As we can see, there is not an evident intent to promote a direct and formal comparison between retrieval practice and deep processing strategies, and as pointed out above, there are

many problems in the discussed comparisons. Furthermore, there must be many similarities and dissimilarities between the common advantages encountered in deep processing and retrieval practice. Even though evidence does not support the elaborative encoding explanation for the testing effect (Karpicke & Blunt, 2011; Lechuga et al, 2015), the allocation of attentional resources and processing effort may be shared mechanisms and occur during study or restudy with deep processing and retrieval practice. Effortful tasks, such as the semantic strategy of Morris et al. (2005) and conceptual mapping, might involve keeping the information in working memory, processing meaning and association, building relationships, and manipulating information. Retrieval practice involves retrieving information, cues, relationships within the material, meaning, and possibly implementing tasks such as those of deep processing. Hence, both can involve effortful processing and attentional resources. Also, both can involve context reinstatement and updating.

There are five main explanations for the retrieval practice effect: retrieval effort, transfer appropriate processing, attentional precedence, elaborative processing, and the episodic-context account. The retrieval effort hypothesis proposes that it is the difficult but successful retrieval of the information that improves the strength of the memory traces and will engender the retrieval practice benefit (Pyc & Rawson, 2009); the second hypothesis is based on the similarity of the testing phase with the final test format, which would facilitate the final performance (Morris, Bransford & Franks, 1977); the attentional precedence is related to the idea that retrieval practice automatically kidnaps or takes precedence over attentional resources (Craik et al, 1996; Buchin & Mulligan, 2017); the elaborative processing hypothesis is that the gain is the result of the use of an elaborative retrieval, with better use of mediators during the test phase (Pyc & Rawson, 2010). The retrieval practice effect is explained, in this case, by semantic elaboration during retrieval, which will engender and exclude some cues, optimizing the use of mediators that will allow retrieving the information in a posterior phase. The episodic-context hypothesis is based on context reinstatement and updating (Karpicke, Lehman & Aue, 2014). Each time retrieval practice is engaged, the individual also retrieves contextual information from the study phase and updates the context with which information is related.

If the retrieval effort hypothesis is correct, we can expect a difference between test and restudy with deep processing conditions, with testing outperforming the deep processing. Although some retrieval can occur during the restudy with deep processing, it is not the main

task. Furthermore, the subject has access to the material, which can facilitate the task and discourage retrieving of information. The same will occur if attentional precedence is correct. If retrieval practice takes precedence and other strategies not, retrieval practice might have a better result.

A testing effect is expected if the elaborative retrieval hypothesis is correct – considering elaborative retrieval and elaborative encoding different processes (Carpenter & Yeung, 2017). Again, the same can be postulated for the case of transfer appropriate processing, if one considers the final test more similar to the study phase with retrieval practice than with deep processing, a retrieval practice effect is expected. If the episodic-context account is right, there would be a benefit in retrieving information, because there will be a context reinstatement and update. Some of those outcomes, as discussed below, can also be different.

Even though retrieval practice is prominent, restudy with deep processing strategies would be better than restudy with shallow processing strategies, because in the deep processing condition the subject has the well-known processing level advantage (Craik & Lockhart, 1972) – some retrieval may also occur while the subject is restudying. Some interpretations can be made differently, one can argue that deep processing strategies can produce the same effort as testing and that the attentional precedence can be applied to deep processing too, which would change the expected results. Also, elaborative encoding and retrieval may be very similar processes with different mechanisms related, for example, to attention or processing effort. When it comes to the episodic-context account, context reinstatement and updating can also occur in deep processing. Another possibility is that both strategies can have similar results that can be explained by different mechanisms. A mixed model, in which more than one of these hypotheses is responsible for the testing effect, is also a discussed outcome. The same can occur in the case of deep processing, with attentional precedence, elaborative processing, and episodic-context account being possible mechanisms.

Thus, here we investigated whether retrieval practice produces greater benefits for memory retention than restudy strategies known to be more powerful than the restudy strategies typically used in retrieval practice experiments. That is, we compared memory retention after 24 hours for studied word pairs that were restudied through shallow encoding tasks, restudied through deep encoding tasks, or subjected to retrieval practice. Based on the findings of the vast majority of retrieval practice studies, we hypothesized that retrieval practice would produce

greater retention than the shallow restudy condition. Because prior research comparing retrieval practice with deep processing restudy is inconsistent both methodologically and in terms of their findings, it is somewhat difficult to predict the outcome for the current comparison between retrieval practice and deep restudy.

2. Experiment 1

In Experiment 1 we investigated the impact of retrieval practice versus deep and shallow processing restudy in a final test using word-pairs as the material. The final test consisted of a recognition task and a judgment of association. Experiment 1 had a within-subject configuration and consisted of two phases. An initial study phase followed by a phase involving three learning strategies: comparing the quantity of consonants (restudy with shallow processing); an imagery task (restudy with deep processing), and cued recall (retrieval practice).

2.1 Methods

2.1.1 Participants

A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample of 37 participants would be necessary to detect the typical medium effect size found for the testing effect, as reported in a recent metanalysis (d = 0.61; Adesope et al., 2017), with α = 0.05 and power of 0.95. To have an equal number of participants in each order/condition arrangement, we went slightly over the suggested sample and recruited 44 participants. All participants were college students (20 women, mean age = 23.07 years, SD = 3.57, range 18-35), were Brazilian Portuguese speakers, and participated voluntarily after signing a consent form attesting their agreement to participate. They were tested individually in sessions lasting approximately 30 minutes on the first day, and 20 minutes on the second day. The study was approved by the Institutional Review Board of the Federal University of Minas Gerais, Brazil (CAAE: 39898514.1.0000.5149).

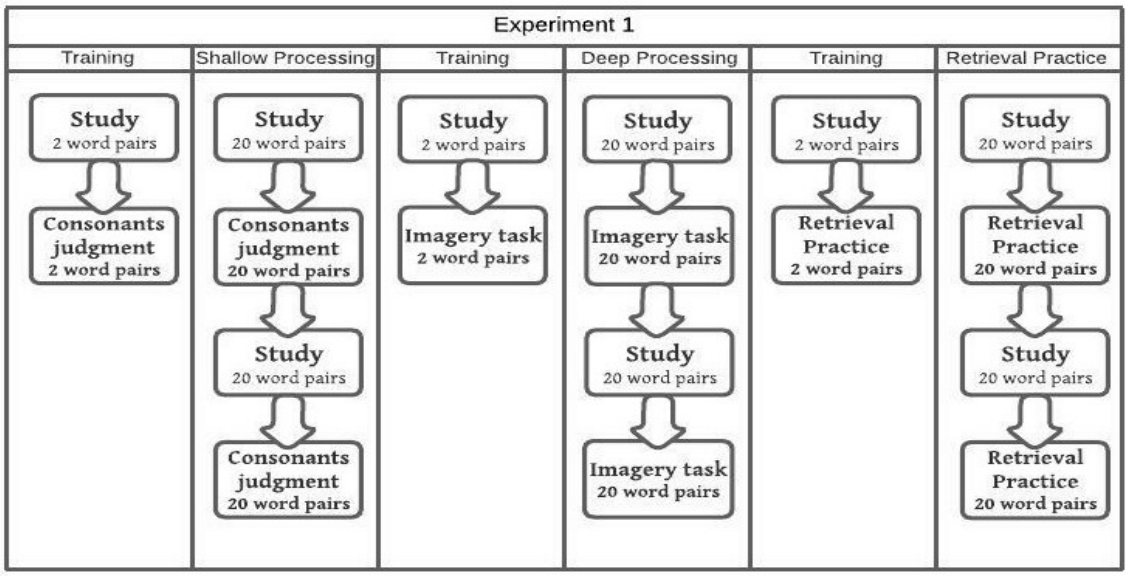
2.1.2 Materials

For each participant, 320 words were randomly selected from a pool of 414 Brazilian Portuguese words with a median frequency of 163 per million and mean concreteness of 2.18 (SD = 0.59) (Janczura et al., 2007). From these selected words, 160 word pairs were randomly matched for each participant; 120 pairs were assigned as targets and presented in both the study and test phases, and 40 pairs were assigned as lures and presented at the final test only. From the 120 target pairs, 60 were rearranged for the recognition task in a manner that their second word (the word on the right of the screen) was replaced by the second word of another rearranged pair. Thus, the second word of all 60 rearranged pairs was randomly relocated within their condition, resulting in new pairs, with no repetition of prior pairs for this condition. All word pairs were shown on the computer screen over a black background. Stimuli presentation and response recordings were performed using PsychoPy v3.0.7 (Peirce, 2007).

2.1.3 Procedures

The task comprised a study and a test phase separated by an interval of approximately 24 hours (range = 22-26 hours). The study phase contained 6 blocks (each containing 20 word-pairs), comprising 2 ‘Restudy Shallow Processing’, 2 ‘Restudy Deep Processing’, and 2 ‘Retrieval Practice’ blocks. The blocks of the same type were presented contiguously, although the order among the different types of blocks was counterbalanced among participants (i.e., there were a total of 6 different block orders, with approximately 7 participants assigned to each order). Each of the six blocks of the study phase was divided into two parts. In the first part, which was identical for all experimental conditions, participants responded using the keyboard keys whether they could think of a meaningful relationship between the words of each word pair (yes = “n”, no = “m”). Each word pair was presented for 5 seconds in the center of the computer screen over a black background (Arial, white), and during this period participants should produce their responses. After this, the screen was blank for 500ms, and then the following word pair was shown for another 5 seconds.

Figure 1
One of the versions of the first phase - Experiment 1.



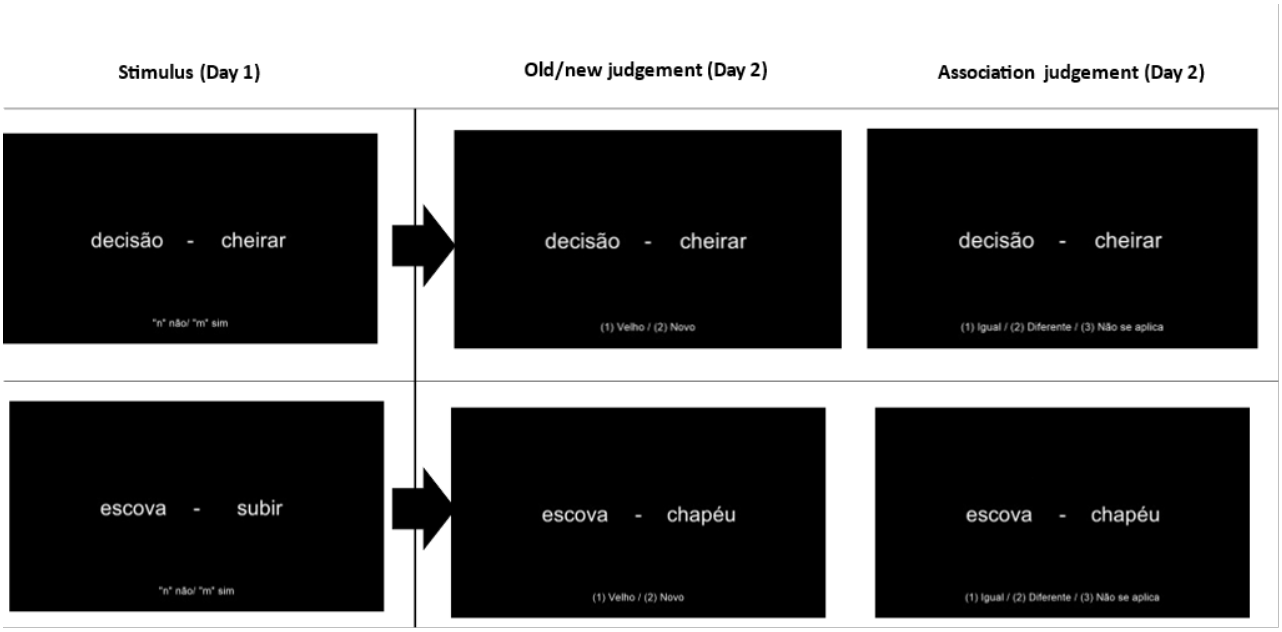
In the second part of each study block, which began immediately after the first part, the same word pairs were presented again for 5 seconds each, separated by blank screen intervals of 500 ms. As can be seen in Figure 1, the task at this part was specific for each experimental condition. Specifically, for the restudy shallow processing condition, participants were asked to judge which of the two words of the pair had more consonant letters: the word on the left (pressing "n"), on the right (pressing "m"), or if they had the same amount (pressing "space"). For the restudy deep processing condition, participants were asked to imagine a scene involving the words of the pair. Finally, for the retrieval practice condition, while the left side word from each pair was fully presented, only the first two letters of the right side word were presented, and participants verbally report the complete word that was previously seen on the right side of the pair. Immediately before the first block of each two types of block, there was a practice phase with two pairs that were not included in the analysis - and with words that were not part of the selected word list (see Figure 1).

At the test, which was about 24 hours after study (range = 22-26 hours), half of the 120 studied pairs were kept intact, and the other half was rearranged. That is, from the 20 word pairs from each study block, 10 were kept intact at the test, and 10 were rearranged. As illustrated in Figure 2, the pair rearrangement consisted of exchanging the right side word of each pair for a word from another pair, from the same study condition (e.g., right side words from the deep processing study condition). All the rearranged and intact pairs were presented in a random order,

intermixed with 40 new pairs (lures). The participant's task was to first judge whether each pair was composed of "old" or "new" words - pressing "1" and "2", respectively. Immediately after each "old" response, participants had to judge whether the pair was intact ("same"), rearranged ("different"), or whether they remembered that the words composing the pair were "old", but do not remember whether they were presented together in the first phase (i.e., whether they were part of the same pair at study). They pressed the 1, 2, and 3 keys to each of these responses, respectively. For the pairs initially classified as "new", no further response was required, and participants were instructed to just press "3" in the second judgment. Before the presentation of each pair, a blank screen was presented for 500 ms. All responses were self-paced, although participants were instructed to respond as quickly as possible without sacrificing accuracy. The subjects knew that some word pairs were rearranged, that each word always appeared on the same side, and that each pair comprised either "old" or "new" words only.

Figure 2.

Recognition task for old/new and association judgment.



Note: The superior line shows an example of a pair that was not rearranged. The inferior line is an example of a rearranged pair.

2.2 Results

To analyze performance at the retrieval practice task on the first day, one point was attributed to each correct response. Correct responses consisted of responding to the word *verbatim*, with even small changes, such as changing the gender or number, classified as an error. Thus, for each participant, the sum of correct responses for each condition was divided by the total number of items from that same condition. We then calculated mean performance for all subjects on the retrieval practice condition at the study phase (first day) using these percentages (Mean= 0.67 SD=0.15).

To estimate recognition performance on the second day, hits were computed by dividing the number of studied word pairs from each experimental condition that were classified as "old" by the total number of studied word pairs from the same condition, whereas correct rejections were computed as the number of unstudied words pairs classified as "new" divided by the total number of unstudied word pairs. For computing the proportion of correct responses for the judgment of association task, we considered correct the responses "different" and "same" for the rearranged and non-rearranged pairs, respectively. Then, the total number of correct judgments made for each condition was simply divided by the total number of pairs from each condition that were previously recognized and judged.

The mean proportion correct for the recognition and judgment of association is presented in table 1. To analyze these data, we conducted two one-way repeated measures ANOVAs with the experimental condition as factors (restudy deep processing, restudy shallow processing, and retrieval practice conditions), one ANOVA was conducted on the recognition data (excluding correct rejections) and the other in the judgment of association data. Finally, we conducted pairwise t-tests to characterize any potential omnibus effect yielded by the ANOVAs.

Table 1

Conditions	**Old/new**	**Judgment of Association**
R. Shallow Processing	0.73(±0.14)	0.72(±0.11)
R. Deep Processing	0.85(±0.10)	0.84(±0.09)
Retrieval Practice	0.85(±0.11)	0.77(±0.08)
New	**0.74(±0.14)**	**-**

Note. The proportion of correct responses for each judgment. Association (relative) was established using the (Total hits/ Total judgments made).

The ANOVA on the recognition data yielded a significant effect, $F(2, 129) = 14.59$, $p < 0.001$, $\eta_p^2 = 0.18$, and the t-tests revealed that retrieval practice produced greater recognition performance than the shallow restudy condition, $t(86) = 4.39$, $p < 0.001$, $d = 0.95$, but equivalent performance relative to the deep restudy condition, $t(86) = 0.12$, $p = 0.902$, $d = 0.01$. As expected, the deep condition yielded greater recognition than the shallow condition, $t(86) = 4.60$, $p < 0.001$, $d = 0.98$. The ANOVA on the judgment of association data also yielded a significant effect, $F(2, 129) = 18.80$, $p < 0.001$, $\eta_p^2 = 0.23$, and the t-tests revealed a significant difference between shallow restudy and retrieval practice, $t(86) = 2.37$, $p = 0.019$, $d = 0.52$ and a difference between deep restudy and retrieval practice, $t(86) = 4.00$, $p < 0.001$, $d = 0.82$. As for recognition, deep restudy yielded greater performance than shallow restudy, $t(86) = 5.76$, $p < 0.001$, $d = 1.19$. Thus, retrieval practice produced a worse performance than the deep processing strategy, and as expected, both produced greater performances than shallow restudy.

Taken together, these results suggest that retrieval practice is worse than restudy when participants are asked to perform a task engaging deep semantic processing during restudy. It raises the question of whether the typically enhanced memory retention reported after retrieval practice is perhaps mainly caused by the frequent adoption of "weak" encoding tasks as a control condition, as rereading, for example. To confirm these patterns, we conducted a second experiment in which the deep encoding restudy task is modified. That is, the imagery task of the current experiment is replaced by a self-reference task.

3. Experiment 2

We made three modifications relative to Experiment 1, which were in the first and second phases. The first was increasing the number of blocks - 4 subsequent blocks for each condition – but maintaining the number of pairs. The aim was to increase the performance in the first phase for the retrieval practice condition. The second modification was the strategy of the restudy deep processing condition: we chose a self-referenced task, in which the subjects had to think of a self-centered situation, event, or story that contained both words of each pair. The third change was in the number of foils, we added 20 foils, ending up with 60 – this change was made because we split the blocks, which could have increased the general performance in the old/new judgment.

The choice of the self-referenced task was based on four aspects: not all materials or words allow good mental imagery processing and the subjects could only think of situations or events in which the information can be applied, associated, related to past memories, or they could have constructed phrases; it increases the relevance, coherence, and applicability of information; the subject uses its semantic memory system to engender a relationship between the words; it maintained the benefit of context reinstatement and updating.

3.1 Methods

3.1.1 Participants

A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample of 17 participants would be necessary to detect the large effect size found in Experiment 1 (d = 0.95), with $\alpha = 0.05$ and power of 0.95. To ensure we had enough power even if slightly different effect-size for the testing effect was produced, we recruited 23 participants. The participants were undergraduate students (12 women, mean age = 22.8, SD =3.04, range = 18-35), Brazilian Portuguese speakers, and participated voluntarily after signing a consent form attesting their agreement to participate. They were allocated to one of the six versions of the experiment, ending up with approximately 4 students for each version. They were tested individually in sessions lasting approximately 30 minutes on the first day, and 20 minutes on the second day. The study was approved by the Institutional Review Board of the Federal University of Minas Gerais, Brazil (CAAE: 39898514.1.0000.5149).

3.1.2 Material

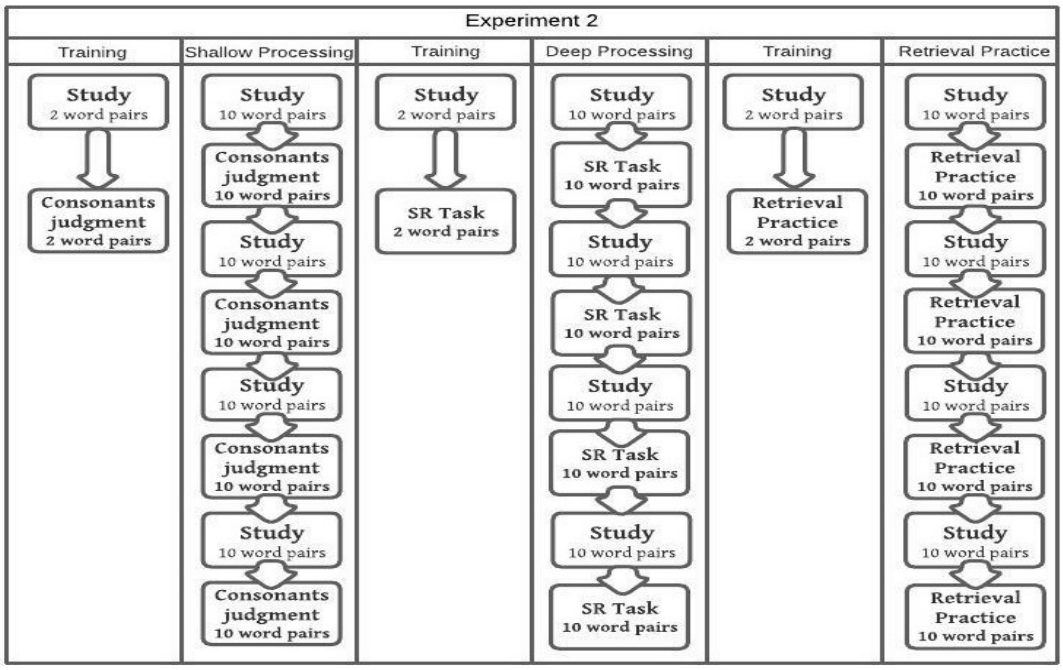
We used the same list of words of experiment 1. The difference was that we selected 180 randomized word pairs, 120 for the study phase and 60 for the second phase of the experiment to be used as foils. We used the software PsychoPy v3.0.7 to build all six versions of the experiment.

3.1.3 Procedures

In contrast to Experiment 1, the experiment was divided into twelve blocks, with four subsequent blocks for each condition, with 10 word pairs that were studied using the same task of Experiment 1 in the retrieval practice and restudy shallow processing. In the restudy deep processing, as explained, there was a self-referenced thought task. This task was thinking of an event, situation, or story in which they were the main characters and contained both words of the pair, they had five seconds to think of the situation, so we oriented them to produce a simple situation. The training phase was the same as Experiment 1, except that after ending the second part of the training of the restudy deep processing condition we requested the subjects to describe the situations that they thought, which was introduced due to the complexity of the self-reference task. After guaranteeing that the subject understood the task, the experimenter also gave two examples of situations – one unreal and one real. For example, for the word pair "neighborhood" and "maintenance" the experimenter always described a real situation: "I am arriving at home and the electric company is doing the maintenance in the electric network of my neighborhood". For the word pair "force" and "antenna", the experimenter always described an unreal situation: "An alien force is trying to influence me using my antenna". During the experiment phase – which occurred immediately after the training, they did not have to describe the situation to the experimenter. The test phase was conducted approximately 24 hours (range = 20-28 hours) after the study phase, and was identical to that of Experiment 1, except that 20 new word pairs were added as foils.

Figure 3

One of the versions of the first phase - Experiment 2.



3.2. Results

The procedures to score performance at retrieval practice, as well as to score the proportion of correct responses for the recognition and judgment of association tasks, were identical to those of Experiment 1. Performance during retrieval practice, on the first day of the experiment, was of 77% correct (Mean = 0.77, SD=0.11), a result that indicates that the reduction in the number of items tested in each block benefited performance at this phase. The remaining data analysis was identical to those of Experimental 1.

Table 2

Learning conditions	Recognition	Judgment of association
R. Shallow Processing	0.76(±0.13)	0.70(±0.10)
R. Deep Processing	0.88(±0.10)	0.80(±0.09)
Retrieval Practice	0.89(±0.07)	0.75(±0.12)
New	**0.69(±0.16)**	**-**

Note: hits percentage in each judgment. Association(relative) was established using the (Total hits/ Total judgments made).

The ANOVA on hits revealed a significant difference between the three conditions, $F(2,66) = 11.74$, $p < 0.001$, $\eta_p^2 = 0.26$, and the t-tests revealed significant greater performance for retrieval practice relative to shallow restudy, $t(44) = 4.27$, $p < 0.001$, $d = 1.25$, greater performance for deep restudy relative to shallow restudy, $t(44) = 3.52$, $p = 0.001$, $d = 1.03$, but equivalent performances for retrieval practice and deep restudy, $t(44) = 0.48$, $p = 0.635$, $d = 0.12$. The ANOVA on the judgment of association data also revealed a significant difference between conditions, $F(2,66) = 5.19$, $p = 0.008$, $\eta_p^2 = 0.14$, and the t-tests showed no significant difference in performance for retrieval practice than shallow restudy, $t(44) = 1.58$, $p = 0.122$, $d = 0.45$. Importantly, however, deep restudy produced greater performance than shallow restudy $t(44) = 3.42$, $p = 0.001$, $d = 1.05$. No difference was found between deep restudy and retrieval practice $t(44) = 1.55$, $p = 0.127$, $d = 0.47$. Thus, further than reinforcing the finding that deep restudy can be as powerful to boost memory as retrieval practice, we show here that deep restudy can produce greater retention, when compared to shallow restudy.

4. General discussion

In both experiments, restudy deep processing produced a very similar result in the recognition task when compared to the retrieval practice condition, resulting in no testing effect between them. The results also showed that when it comes to associative memory, deep processing strategies may produce a better outcome. Restudy with shallow processing, on the other hand, obtained the worse result in almost all measures. Although it was expected, this reinforces that the result can't only be explained by a good study phase, which involved a deep processing strategy.

Using the visual imagery strategy improved the memory for the association of the word pairs in absolute, the subjects remembered a bigger quantity of items' association when they used the deep processing strategy and were more accurate. When it comes to the self-referenced task, the subjects were also more accurate in the judgment of association when compared to shallow restudy – but not retrieval practice. Retrieval Practice didn't improve the performance in associative memory in the second experiment.

The fact that there was no difference between restudy DP and testing in the recognition task culminate in at least three possible ways to explain why these different strategies had a

similar performance: a shared mechanism, different mechanisms that are equally effective for a posterior old/new judgment and mixed mechanisms that partially explain the results, but that allows a flexible explanation in the case of the judgment of association.

The results are consistent with some of the hypotheses that are postulated with the intent to explain the retrieval practice effect. One could purpose that the better results obtained with the visual imagery task, self-referenced task, and retrieval practice can be explained by the retrieval effort hypothesis, the attentional precedence hypothesis, elaborative processing, or the episodic-context account.

Retrieval effort can explain the performance in the retrieval practice condition, but not the restudy deep processing. Buchin & Mulligan (2017) reject it because retrieval practice was less jeopardized in their experiment than would be expected if it was the case. Also, restudy deep processing was at least as good as retrieval practice and retrieval isn't an automatic or obligatory process during restudy. In this case, their performances would be explained by two different mechanisms. Another possibility is to consider a mixed model in which the deep processing strategy is equivalent or similar to retrieval practice when it comes to processing effort, possibly combining the attentional precedence and the effort related, for example, in retrieval and/or elaborative processing.

When we compare deep processing strategies and retrieval practice, both probably recruit attentional resources when the subjects are engaged in processing or retrieving the information. Deep processing also promotes elaborative processing and/or context reinstatement and update that allows the establishment of mediators and contextual information that helps in the final recollection. The subjects must establish a relationship between items, culminating in an association, which is not always automatic, and this possibly generates a benefit in the final judgment of the association task. Also, the visual imagery task and the self-referenced task may produce a dual code (Paivio & Lambert, 1981).

Both retrieval practice and restudy shallow processing involve the effortful processing of information. One can argue that making the consonant comparison between the two words of the pair certainly requires effort and attention. But, on the other side, they are not allocated in the processing of meaning or for retrieving information. This important difference helps to explain the difference between these strategies and the restudy SP, but not exactly the difference between the deep processing strategy and retrieval practice.

Taking the deep and shallow processing, it is expected that deep processing strategies would promote better performance in a final test when compared with shallow processing strategies. An extensive amount of data helps to explain this phenomenon. Levels of Processing per se, transfer appropriate processing – in our case, attentional processes and a possible transient task that impedes adequate processing - in the level of meaning, like a judgment of consonants – might explain the difference between the two conditions. The shallow processing strategy may have acted as an undesirable difficulty.

The elaborative processing is at the root of the deep processing and it can explain at least partially the evidence that was found in this experiment – the preponderance of restudy DP over restudy SP. But the similarity of restudy DP and retrieval practice conditions in the performance in the recognition task and the concomitant difference in associative memory is inconsistent with the elaborative explanation. It wasn't only the elaborative processing that generated the advantage of both strategies over restudy SP.

First, having no benefit in retrieval practice contrasts the findings in Karpicke & Blunt (2011), Karpicke & Smith (2012), and Neuschatz et al (2005) that showed a testing effect using conceptual maps and name-face imagery, respectively. As already mentioned, they also contrast because there was a significant difference between restudy DP and retrieval practice in the judgment of association task. If the benefit was explained by elaborative encoding/retrieval, they shouldn't have a similar result in the recognition task and a different result in the judgment of association. Third, the elaborative hypothesis was already questioned or rejected in other studies (Buchin & Mulligan, 2017; Buchin & Mulligan, 2019; Mulligan & Picklesimer, 2016)

One can argue that the results in the recognition task and the lack of difference in the performance in the judgment of association of Experiment 2 don't allow the immediate rejection of this hypothesis, but even some defenders of this hypothesis can reject it by considering, as Carpenter & Yeung (2017), elaborative retrieval different from elaborative encoding. In this case, the expectation would be the opposite: a higher performance in the judgment of association for retrieval practice. It is possible to purpose, on the other hand, that both mechanisms improve memory performance in two different ways. We postulate that, although retrieval practice may allow elaborative processing, it is not an obligatory or automatic strategy during the retrieval phase, it can be only one of the processes that help to explain the benefits - when it is

implemented. Furthermore, if elaborative processing was more prominent only in the restudy DP, there would be a difference in performance in the recognition task.

General conclusions about the transfer-appropriate processing hypothesis cannot be consistently made because the task in the retrieval practice condition was different in some way from the final test, but some propositions are possible. In the recognition task, the subjects needed to recollect and/or be familiar with only one of the items of the pair. In our experiment, retrieving one item didn't give any advantage in the recognition task when compared with restudying the items with deep processing. The obligatory construction of a relationship in the visual imagery task and self-referenced thought may be more like the judgment of association task than the retrieval practice, but at the same time, retrieval practice usually gives an advantage for recollecting information in the final test, if the subjects had a better recollection in the second day, they would also remember the pairs' association, but our final result contrasts this expectation. The similarity in the recognition task, with both having a higher performance than restudy with shallow processing, plus the lack of difference in the judgment of association in the first experiment and the superiority of restudy deep processing in the second experiment impose difficulty in explaining those results with the transfer appropriate processing.

The episodic-context account can explain both retrieval practice and restudy deep processing results. Karpicke, Lehman & Aue (2014) criticize and differentiates the elaborative hypothesis from the episodic-context account and some principles related to the episodic-context account can be applied to restudy with deep processing: the imagery task involves building a different context, it is also possible that some retrieval and manipulation of information – which can involve reinstating the previous context – might occur during the task. Furthermore, the context during the restudy with deep processing tends to be more different than the study phase - when compared, for example, to the retrieval practice condition. This would update the context and limit the search field for the final test. A similar process occurs during the retrieval practice. The authors also discuss how some difficulties are not desirable and may harm the memory performance in a posterior test. This helps to explain why the consonant's judgment obtained the worse result, although it didn't involve retrieval practice, the task, as postulated above, can be an undesirable difficulty.

We advocate in favor of attentional precedence and episodic-context account because the deep processing strategy and retrieval practice probably demand more attentional resources and

allow context reinstatement and updating. As discussed, the effort imposed by the shallow processing task and the cued recall didn't promote any advantage when compared with the deep processing task. The fact that the episodic-context account is built upon the preponderance of retrieval practice didn't exclude that the same advantage, related to the contextual information, can be applied to restudy DP. Deep processing strategies engender much contextual information, mediators, and cues. The idea that during the retrieval practice, the imagery task and the self-referenced thought task there is a retaking of the original context, that appears mixed to the new context that is built, and that this will lead to a context update is valuable to explain our data. Even the difference between retrieval practice and the deep processing task in associative memory can be explained by the desirable difficulty added by the imagery-task – in which there is a conversion of verbal to non-verbal information. The last also limit the search field for the studied items – even more than the imagery task – because only items that were processed in the non-verbal context might be remembered. It also explains why the difference between retrieval practice and deep processing when compared to shallow restudy emerged after the improvement of performance on the first day: it reduced the desirable difficulty related to the distance – in terms of time and context- between study and retrieval practice.

The attentional precedence permits to explain the better performance in the recognition task of both retrieval practice and restudy DP. The more dedicated context reinstatement and updating, which is not demanded in the cued recall as it is in the imagery and self-referenced task, helps to explain the difference in accuracy and the number of pairs remembered between the two conditions. Hence, the attentional precedence hypothesis doesn't explain the process that underlies the advantage. The combination of the allocation of attentional resources, processing effort, and contextual reinstatement and updating helps to explain the results of deep processing. The last one is less present in the retrieval practice condition of our experiment, which involved more of the attentional precedence and retrieval effort.

Most of the studies that compared retrieval practice with possibly good and strong strategies were made with configurations significantly different. Some differences were: using a between subjects' paradigm, the kind of material that was studied, the lack of assessment of associative memory, the level of exposure to the material, and strategies that depend highly on the experimenter's or the subjects' level of dexterity.

Our findings go along with the ones of Fritz et al. (2007) that found a similar result between retrieval practice and the keyword mnemonic – a strategy that involves visual imagery. Also, with the absence of testing effect found in some measures in McDaniel et al. (2009). Some differences from our study can be pointed: the study of Fritz et al (2007) used a longer interval – 20 seconds- to produce an image. The authors matched this time in the retrieval practice condition by giving the subjects more retrieving opportunities. We got no testing effect with only five seconds to produce a mental image or a self-centered situation.

Fritz et al. (2007) used an arbitrary relationship - such as Neuschatz et al (2005)- between words of two languages, in ours, both words were Portuguese, this is important because the connection that was established by the subjects in our experiment involved semantic processing and not only pure correspondence. These dissimilarities need to be emphasized because associative memory is very important for connecting information using semantic proximity -and not necessarily arbitrary relations.

An important finding of Fritz et al. (2007) was, alternatively, having no difference in performance even when the subjects had many opportunities to retrieve information. This reinforces that there may be other strategies that are at least as good as retrieval practice. The study of Neuschatz et al. (2005) had a preponderance of the retrieval practice.

The focus in these studies is related to their proximity with ours, as discussed, a vast literature shows the advantage of retrieval practice when compared with restudy (Einsenkramer & Jaeger, 2013; Rowland, 2014) and some show a benefit even with possible strong strategies (Karpicke & Blunt, 2011; Morris et al., 2005; Neuschatz et al,2005; Rummel et al., 2017; Stenlund et al., 2017).

Studies using conceptual mapping, which can be considered a deep processing strategy found a consistent testing effect (Karpicke & Blunt, 2011; Lechuga et al., 2015; O'Day and Karpicke, 2020). Conceptual mapping, as discussed, is a technique that requires some training or expertise (Lechuga et al., 2015). Furthermore, conceptual mapping is very different from our strategies, it may reduce the cognitive load because the subject can write and reread the information to-be-remembered, allowing the subject only to study or restudy the material and the relationships contained in the material.

Both hypotheses that might help to explain our results don't completely apply to this last strategy: processing effort and attentional precedence. This leaves conceptual mapping with the

elaborative encoding advantage that is also questionable because the text can give or facilitate the association of the items. It is also probable that the subjects spend their time restudying or even not doing anything while they have the material and the paper to build their maps. The last problem is reported in the next study that we will discuss.

Stenlund et al. (2017) compared the retrieval practice with feedback with group discussions. Even the condition of group discussions with feedback got a worse result when compared to retrieval practice. The authors argue that group discussions may act like restudy and demonstrated that some individual differences mediate the performance in the group discussions. They also showed that subjects tend not to use all the time for practicing with the group.

When it comes to the applicability, although not all the materials can be easily manipulated using visual imagery, many have some correspondence with non-verbal material, images, or can be adapted to this end. Another possibility is creating usual images that can be reproduced during the courses or at least give the student a way to start building an image. The teacher's and educational institutions' experience and cumulative knowledge will allow the creation of imaging prototypes. Abstract information can be adapted to imaging or self-reference thought by bringing the discussions and texts to real-world examples.

Self-referenced thought may be an interesting strategy in the educational context because it first demands and allows the subjects to process the information at the semantic level; the information can be related to previous knowledge and, even if the subjects didn’t have any previous information that would help them, they can create a situation in which the material studied can be applied or processed using different perceptual systems.

Visual imagery is not always possible - mainly with complex and abstract material, but self-referenced thought is more adaptable: the subject will only bring the information to the way or to the context in which he or she lives. It also may induce proximity of the information to-be-learned and the life per se, which can affect the capability of recollecting the applicability and generalizing the use of information. Lastly, it allows the subject to build a significantly distinct context, which can improve the performance in a final test. Unfortunately, not all hypotheses could be tested with the material and the paradigm that we used.

Strategies such as keyword mnemonic, conceptual mapping, and name-face imagery are very difficult to apply without training, require too much time from the student, and may need substantial support from a tutor or teacher, because there may be specific information that is used

during the study or restudy phase. Group discussions take too much time and are not well suited for all the students (Stenlund et al., 2017)

Using deep processing strategies can also involve a retrieval practice, because the subjects may stop studying the material and try to access the presented information. Far from being a problem, as shown in Collins et al. (2018) it may produce a better result than retrieval practice with shallow processing strategies. So even combining these strategies may be beneficial.

A different way to combine these strategies is using a sequential approach, in this sense, subjects will be exposed to both strategies in two different moments. One can even use three study phases – using pre-testing in the first phase (Lima & Jaeger, 2020; Latimier et al, 2019). The results found in Karpicke & Smith (2012) and O'Day & Karpicke (2020) showed a worse result when combining the strategies. Fritz et al (2007) also compared the combination of strategies with the strategies alone - using retrieval practice and keyword mnemonic - and they didn't find any difference. Neuschatz et al (2005) found a benefit.

Although the results are controversial, mixing is also almost inevitable - retrieval practice, for example, always requires a study phase. Students commonly use many learning strategies at once (Minear et al., 2018) - not necessarily using the best ones. Choosing those strategies that deliver the best performance can optimize learning, investment, and planning in the educational context. phase. This way, the main question ends up being which strategies we should combine.

One can also question if subjects already use good strategies for studying, this could cause a serious problem because educators might change what is already going well for a weaker strategy. This problem may occur because many strategies can't be empirically compared – for many reasons – and because there may be individual differences in the capacity to benefit from each strategy.

These fears are mitigated by some findings. Dunlosky et al. (2013) showed that many of the most commonly used strategies don't possess strong evidence to support their use. In the study of Fritz et al (2007), when the subjects chose their learning strategy, the result was worse than both retrieval practice and keyword mnemonic. Roediger & Karpicke (2006a) also showed that students weren't good at predicting their final results when restudy and retrieval practice were compared.

This study has some limitations: the first is the use of word pairs, which are not the kind of material that is usually found in the educational context. Their use is justified to generate a

great number of mini-events to investigate the possible effects and the associative memory aspect. The fact that our sample was composed of college students, that it was a laboratory experiment, and that there were inferior and superior age limits also jeopardize the generalizability of our findings.

The partition and use of mini-events also limit the quantity of information that the subject must process while restudying or testing, which also imposes a difficulty for inferences or generalization of our findings. Increasing the amount of information to be remembered may hamper the use of visual imagery or self-referenced thought. Although it is well known that postponing the retrieval practice will also reduce the final performance, the testing effect keeps appearing. If this problem is reproduced for the case of deep processing, one can present a limited amount of information per period – which in our experiment, led to better performance.

A within-subject paradigm always has the problem that the subject can use one strategy in another part of the experiment. An advantage is the possibility of mitigating the effect of individual differences when one compares two or more strategies. Furthermore, it allows obtaining a medium performance of many strategies using the same subject. Exposing the same subject to two different strategies may be a source of comparison in individual differences in the benefit that each technique can produce. Finally, our findings are sustained even in a paradigm that could have enhanced retrieval-induced forgetting.

5. Conclusions

In sum, the consistency of our results evidences the power of two deep processing strategies when compared with the retrieval practice – one of the most powerful and studied learning strategies. Future studies should explore if these benefits are maintained with different materials, with different kinds of test formats, and with different age ranges. The hypotheses here postulated should also be tested to understand the underlying mechanisms that generated the advantages of deep processing strategies and retrieval practice, and lastly, the superiority in the judgment of association for the deep processing strategy.

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(3), 659-701. https://doi.org/10.3102/0034654316689306

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, 24(1), 43–56. https://doi.org/10.1037/xap0000133

Biwer, F., oude Egbrink, M. G., Aalten, P., & de Bruin, A. B. (2020). Fostering Effective Learning Strategies in Higher Education–A Mixed-Methods Study. *Journal of Applied Research in Memory and Cognition*, 9(2), 186-203. https://doi.org/10.1016/j.jarmac.2020.03.004

Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of experimental psychology: learning, memory, and cognition*, *43*(12), 1934 - 1947. doi:10.1037/xlm0000427

Buchin, Z. L., & Mulligan, N. W. (2019). Divided attention and the encoding effects of retrieval. *Quarterly Journal of Experimental Psychology*, 72(10), 2474-2494. https://doi.org/10.1177/1747021819847141

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36 (5), 1118-1133. doi:10.1037/a0019902

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14 (3), 474-478. doi:10.3758/BF03194092

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, 92, 128-141. https://doi.org/10.1016/j.jml.2016.06.008

Chang S. H. (2017). The Effects of Test Trial and Processing Level on Immediate and Delayed Retention. *Europe's journal of psychology*, 13(1), 129–142. https://doi.org/10.5964/ejop.v13i1.1131

Collins, D. P., Rasco, D., & Benassi, V. A. (2018). Test-enhanced learning: Does deeper processing on quizzes benefit exam performance?. *Teaching of Psychology*, 45(3), 235-238. https://doi.org/10.1177/0098628318779262

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

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. doi: 10.1177/1529100612453266

Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology* (Ruger H.A, & Bussenius C.E, trans). Teachers College, Columbia University, New York. (Original work published in 1885).

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. doi:10.1590/1982-43272356201314

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191. https://doi.org/10.3758/BF03193146

Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, 21(4), 499-526. https://doi.org/10.1002/acp.1287

Janczura, G. A., Castilho, G. M. D., Rocha, N. O., Van Erven, T. D. J. C., & Huang, T. P. (2007). Concreteness norms for 909 portuguese words. *Psicologia: teoria e pesquisa*, *23*(2), 195-204. doi:10.1590/S0102-37722007000200010

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772-775. doi:10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B.H Ross (Ed.), *Psychology of learning and motivation,* 61, 237-284. Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Latimier, A., Riegert, A., Peyre, H., Ly, S. T., Casati, R., & Ramus, F. (2019). Does pre-testing promote better retention than post-testing?. *NPJ science of learning*, 4(1), 1-7. https://doi.org/10.1038/s41539-019-0053-1

Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. Learning and Instruction, 40, 61-68. https://doi.org/10.1016/j.learninstruc.2015.08.002

de Lima, N. K., & Jaeger, A. (2020). The Effects of Prequestions versus Postquestions on Memory Retention in Children. *Journal of Applied Research in Memory and Cognition*, *9*(4), 555-563. https://doi.org/10.1016/j.jarmac.2020.08.005

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, 20(4), 516-522. https://doi.org/10.1111/j.1467-9280.2009.02325.x

McDermott, K. B. (2021). Practicing Retrieval Facilitates Learning. *Annual Review of Psychology*, 72, 609-633. https://doi.org/10.1146/annurev-psych-010419-051019

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533. https://doi.org/10.1016/S0022-5371(77)80016-9

Morris, P. E., Fritz, C. O., Jackson, L., Nichol, E., & Roberts, E. (2005). Strategies for learning proper names: Expanding retrieval practice, meaning and imagery. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(6), 779-798. https://doi.org/10.1002/acp.1115

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 938-950. doi:10.1037/xlm0000227

Neuschatz, J. S., Preston, E. L., Toglia, M. P., & Neuschatz, J. S. (2005). Comparison of the efficacy of two name-learning techniques: Expanding rehearsal and name-face imagery. *The American journal of psychology*, 118(1), 79-102. doi: 10.2307/30039044

O'Day, G. M., & Karpicke, J. D. (2020). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology*, 113(5), 986–997. https://doi.org/10.1037/edu0000486

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of verbal learning and verbal behavior*, 20(5), 532-539. https://doi.org/10.1016/S0022-5371(81)90156-0

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13. doi: 10.1016/j.jneumeth.2006.11.107

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. doi:10.1016/j.jml.2009.01.004

Pyc, M.A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335-335. doi:10.1126/science.1191465

Roediger III, H. L., & Karpicke, J. D. (2006a, March 1). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255. doi:10.1111/j.1467-9280.2006.01693.x

Roediger, H., & Karpicke, J.D. (2006b, September 1). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1, 181 - 210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1463. doi:10.1037/a0037559

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking?. *Journal of Experimental Psychology: Applied*, *23*(3), 293. https://doi.org/10.1037/xap0000134

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641. https://doi.org/10.1037/h0063404

Stenlund, T., Jönsson, F. U., & Jonsson, B. (2017). Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics. *Educational psychology*, 37(2), 145–156. https://doi.org/10.1080/01443410.2016.1143087

Su, N., Buchin, Z. L., & Mulligan, N. W. (2021). Levels of retrieval and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(4), 652–670. https://doi.org/10.1037/xlm0000962

ANEXO A - TCLE

Termo de Consentimento Livre e Esclarecido - TCLE

Título da Pesquisa: *O efeito de testes na retenção de memórias.*

Prezado (a),

Este é um convite para você participar *voluntariamente* em uma pesquisa que irá investigar a sua memória. Estamos à disposição para esclarecer quaisquer dúvidas em relação à pesquisa antes e durante a execução da mesma. Certifique-se que você compreendeu todas as informações abaixo antes de expressar ou não o seu consentimento para participar da pesquisa.

1. Objetivos e justificativa do estudo

O objetivo desta pesquisa é investigar como a prática de evocar memórias pode auxiliar na retenção das mesmas. Isto é, iremos investigar se evocar memórias durante testes é uma boa estratégia para a memorização de longo prazo.

2. Procedimentos da avaliação

Caso você concorde em participar deste experimento, será realizado um teste em que a sua capacidade de memória para pares de palavras será avaliada com um intervalo de uma semana ou um dia. Assim, sua participação no estudo é iniciada hoje, e finalizada dentro de 1 ou 7 dias, ocasião na qual um teste final será realizado.

3. Realização da Pesquisa

A pesquisa está sendo liderada pelo Professor Antônio Jaeger, vinculado ao departamento de Psicologia da Faculdade de Filosofia e Ciências Humanas (FAFICH).

4. Participação voluntária e sem compromisso financeiro

Sua participação é voluntária e não implica em nenhum compromisso financeiro entre você e a equipe da UFMG. Você não terá gastos com a pesquisa, e para evitar que haja gastos com transporte exclusivos para a pesquisa, sua participação será agendada considerando os horários que você já estará presente no Campus Pampulha da UFMG, local onde o experimento ocorrerá.

5. Liberdade de recusa e de desistência

Você poderá negar o consentimento ou mesmo desistir de participar durante qualquer fase da pesquisa, sem nenhum prejuízo e sem necessidade de se justificar.

6. Garantia de sigilo

Os resultados da pesquisa serão utilizados em trabalhos científicos publicados ou apresentados oralmente em congressos e palestras, sem revelar a sua identidade ou quaisquer informações particulares, ou que possam de qualquer maneira identificá-lo(a).

7. Riscos

O maior risco está relacionado principalmente ao cansaço durante a testagem psicológica. Todo esforço será feito no sentido de atentar para o seu bem-estar físico e psicológico, interrompendo-se a testagem aos menores sinais de desconforto.

8. Benefícios em participar da pesquisa

Os benefícios consistem no potencial aprendizado de estratégias de memorização, que se utilizadas de maneira disciplinada pelo participante, poderão ter efeitos benéficos no futuro acadêmico do mesmo. É importante ressaltar que se tratam de estratégias de aprendizagem que estão ainda em fase experimental, e dependerão de confirmações científicas quanto a sua eficácia, assim como do empenho de cada indivíduo, para que vantagens acadêmicas reais no futuro sejam observadas.

Atenciosamente,

_____

Prof. Dr. Antônio Jaeger
Coordenador da Pesquisa
Professor Adjunto do Departamento de Psicologia da UFMG
Av. Antônio Carlos, 6627, FAFICH-UFMG, Sala 4060
Laboratório de Neuropsicologia do Desenvolvimento
Tel: (31)34096295 / E-mail: antonio.jaeger@gmail.com

Para maiores esclarecimentos sobre dúvidas éticas você pode consultar também o Comitê de Ética em Pesquisa (COEP-UFMG), na Av. Antônio Carlos, 6627 – Unidade administrativa II, 2º andar/ Campus Pampulha- UFMG, Tel: (31)34094592/ E-mail: coep@prpq.ufmg.br ou a Comissão Nacional De Ética Em

Pesquisa (CONEP), na SEPN 510 NORTE, BLOCO A 1º subsolo, Edifício Ex-INAN - Unidade II - Ministério da Saúde, Tel:(61) 3315-5878

Eu, _____,
abaixo assinado(a), declaro ter sido informado(a) e esclarecido todas as minhas dúvidas sobre os procedimentos e propostas da pesquisa '*Aspectos episódicos do efeito de testagem*' e concordo em participar voluntariamente da mesma.

_____

Assinatura

Belo Horizonte, _____ de _____de _____

Contato telefônico (Preenchimento não obrigatório): (____) _____

Campo para ser preenchido pelo pesquisador

Informações gerais:

Sexo: _____ Idade: _____ Iniciais: _____.

Número do participante: _____ Experimento: _____.

Grupo do participante no experimento: _____.

Experimentador: _____. Data: _____/_____/_____.

Observações:

_____
_____
_____
_____
_____.