Universidade de Brasília

Institute of Psychology

Department of Basic Psychological Processes

Graduate Program in Behavioral Sciences

Master’s Thesis

The efficacy of retrieval practice in creating stress-resistant memories

Carlos Eduardo Dantas da Costa Lage

Brasília, November 1, 2019.

Universidade de Brasília

Instituto de Psicologia

Departamento de Processos Psicológicos Básicos

Programa de Pós-Graduação em Ciências do Comportamento

Dissertação de Mestrado

A eficácia da prática de lembrar na criação de memórias resistentes ao estresse

Carlos Eduardo Dantas da Costa Lage

Brasília, November 1, 2019.

The efficacy of retrieval practice in creating stress-resistant memories

by

Carlos Eduardo Dantas da Costa Lage

Dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in Behavioral Sciences in the Graduate Program in Behavioral Sciences at the University of Brasília (Research area: Cognition and Behavioral Neurosciences).

Advisor: Prof. Luciano G. Buratto, Ph.D.

Examining Committee

Prof. Dr. Luciano Grüdtner Buratto, Ph.D. (President)

Graduate Program in Behavioral Sciences

University of Brasília (UnB)

Prof. Dr. Antônio Jaeger, Ph.D. (External member)

Graduate Program in Psychology: Cognition and Behavior

Federal University of Minas Gerais (UFMG)

Prof. Dr. Ricardo Moura, Ph.D. (Internal member)

Graduate Program in Behavioral Sciences

University of Brasília (UnB)

Prof. Dr. Ricardo Basso Garcia (Substitute member)

Graduate Program in Behavioral Sciences

Department of Basic Psychological Processes

University of Brasília (UnB)

Brasília, November 1, 2019.

Dedications

This work is dedicated to my parents, Geraldo Magela da Costa Lage and Evelyse Klier Dantas, as well as to my children, Gabriel Macoski Leite Dantas Lage and Carolinna Macoski Leite Dantas Lage.

To Mom and Dad, my deepest gratitude for everything you have done, from loving me unconditionally to instilling responsibility, respect and moral/ethical values in me, helping me grow into the man I have become. By watching and listening to you, I was taught how to face everything and rise, never letting go of my dreams. I thank God for you lives, certain that I could not have asked for better role-models and guides in my life.

To my kids, God granted me with the opportunity to be your father and I have no words to say how gratifying and fulfilling this is to me. Through our coexistence, I have learned much more than I could have ever expected. You are the most important people in my life. I pray that my examples in life guide you to an outstanding existence, like my parents' examples guided me. Believe that dreams do come true, but only if we are perseverant and resilient enough to keep walking towards our goals in life, no matter what. God bless you!

Acknowledgements

Thank God! Throughout the development of this research I have received incommensurable support and assistance. I would first like to deeply thank, Dr. Marcia Schelb and Prof. Dr. Luciano Grüdtner Buratto, two very special people who have been essential on this journey.

I passionately thank Dr. Marcia Schelb, my beloved wife, who has unconditionally supported my decision to pause my executive career development to invest a large amount of time and effort into this academic aspiration, with all the challenges it brought into our lives. You have always encouraged me and provided me with all that was necessary, by maintaining a supportive and affective environment through these 2 laborious years, hence enabling me to focus a great deal of energy on this project. Thanks for your partnership and sympathetic ear. I will be forever grateful for your love.

I seriously thank Prof. Dr. Luciano Grüdtner Buratto, my dear advisor, whose expertise, comments, remarks and partnership were pivotal for designing, conducting and understanding this research. Your guidance, promptness and involvement really played a great role for me on setting the foundation of how to produce scientific relevant experimental research focused on excellence. You opened a door, and walked along with me through the academic path, which led me to a higher educational development and helped me move forward on my personal journey to understand the neurocognitive-behavioral processes involved in learning.

I would like to thank all my memory research group colleagues for the valuable collaboration on exploring scientific relevant literature. My very special thanks to Marcos Lima, my “research partner”, who has helped me in several dimensions of this research. Besides all the bibliography we have shared and discussed, programming the experiment and analyzing the results was easier because of your partnership and willingness in sharing your knowledge about the cognitive and statistical sciences with me. I thank you for all the conversations we had that were related to our projects, they have been of great value to me. I would like to thank my two undergraduate research assistants, Giulia Veiga and Júlia Valle, for their partnership and great job that was done on applying the experimental sessions and managing participants' data, samples and schedules, always focused on excellence. I would also like to thank every participant who voluntarily shared their precious time to attend the experimental sessions.

I would like to thank UPTIME Brasília staff, who have supported me, their executive director, by being more engaged in the general management and taking responsibility for the company results while I was developing this research. I thank the Coordination for the Improvement of Higher Education Personnel (CAPES) that supported my research by awarding me with a scholarship grant. I thank Sabin Institute, the analysis laboratory that supported this research, through their research support division (NAP), by processing more than half of the cortisol saliva samples for free.

Last, but sure not least, I would like to thank the Psychology Institute of University of Brasília (IP-UnB), specially the Graduate Program in Behavioral Sciences (PPG-CdC) and the Laboratory for Psychological Research in Humans (LIPSI). I would like to thank every professional that have directly or indirectly helped me with my academic life and research development. Special thanks to all professors I had the opportunity of having classes with, who have helped me build a much better knowledge of the cognitive neuroscience, thus empowering me to achieve my research goals.

Table of Contents

List of Abbreviations, Acronyms and Symbols.

α	Cronbach's alpha (internal consistency index and significance level)
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
BAI	Beck Anxiety Inventory
β	log odds ratio for multilevel logistic regressions
CG	Control Group
D	Difficulty factor
d	Cohen's *d* (measure of effect size on *t*-tests)
F	*F*-ratio (ANOVA's statistic measure)
HPA	Hypothalamic-pituitary-adrenal
JOL	Judgement of learning
k	Cohen's kappa (inter-rater agreement between judges' measure)
M	Sample mean
η_p^2	Partial eta-squared (ANOVA's effect size measure)
OR	Odds ratio
p	*p*-value
RP	Retrieval Practice
RT	Reaction time
S	Strategy factor
s	Seconds
SD	Standard Deviation
SECPT	Social Evaluated Cold Pressor Test
SE	Standard Error

SEM	Standard Error of the Mean
SG	Stressed Group
SNS	Sympathetic nervous system
SP	Study Practice
t	*t*-statistic
WP	Word Pairs used in the experiment as stimuli
Z	Wald statistic

List of Tables

List of Figures

Abstract

Human memory research has shown that practicing retrieval improves long-term memory, and that stress induced before retrieval impairs memory performance. A recent study by Smith, Floerke, and Thomas (2016) has shown that this detrimental effect of stress on memory does not occur when materials are encoded via retrieval practice. In this study, we sought to replicate this protective effect of retrieval practice under conditions of stress, while controlling for item difficulty. Participants (59 undergraduate men) first studied 40 Swahili-Portuguese word pairs. Half of the word pairs were then repeatedly restudied while the other half were repeatedly retrieval practiced. Half of the pairs in each condition were high in memorability (easier), whereas the other half were low in memorability (harder). Participants were then asked about which learning strategy (restudy vs. retrieval practice) would result in better recall on a subsequent test. One week later, participants returned for a final cued-recall test. Half of the participants underwent a stress-induction protocol (modified SECPT) 25 min before the beginning of the memory test for all 40 word pairs; the other half underwent a control condition. Salivary cortisol and questionnaire responses were used to assess the efficacy of stress induction. Results showed that stress was successfully induced in the stress group. In addition, participants recalled more word pairs when learned via retrieval practice than via restudy, replicating previous findings. Unexpectedly, the retrieval practice effect was larger for easier targets than for harder targets. In addition, most participants expected to recall more items in the retrieval practice condition than in the restudy condition, in line with actual performance, but unlike previous results indicating poor metacognitive calibration in similar samples. The protective effect of retrieval practice, as reported by Smith et al (2016), was not replicated, possibly due to a floor effect in the restudy condition. The results, however, revealed an unanticipated interaction in the retrieval practice condition such that stress increased recall for harder items but decreased recall for easier items. We provide a tentative account of this finding by suggesting that retrieval practice of harder items adds a stress-related attribute to the item's memory trace, whose retrieval is later facilitated under stress (conditionalized analysis). The results suggest that the impact of stress on memory depends both on the learning strategy and on the intrinsic difficulty of the material.

Keywords: retrieval practice; cued recall; stress; cortisol; item difficulty.

Resumo

Pesquisas sobre memória humana mostram que a prática de lembrar melhora a memória de longo prazo, e que a indução de estresse antes da evocação prejudica a performance da memória. Um estudo recente de Smith, Floerke, and Thomas (2016) mostrou que esse efeito deletério do estresse não acontece quando o conteúdo foi codificado via prática de lembrar. Neste estudo, nós buscamos replicar esse efeito protetivo da prática de lembrar sob estresse, controlando a dificuldade do item. Os participantes (59 alunos de graduação) estudaram 40 pares de palavras Suaíli-Português. Metade dos pares foram, então, estudados por releitura, enquanto a outra metade foi estudada por prática de lembrar. Metade dos pares em cada condição tinha memorabilidade alta (fáceis), enquanto a outra metade tinha memorabilidade baixa (difíceis). Os participantes foram questionados, ao final da sessão, sobre qual estratégia de estudo (releitura vs. prática de lembrar) traria melhores resultados em um teste subsequente. Uma semana depois, os participantes retornaram para o teste final de memória. Metade dos participantes foi submetida a um protocolo de indução de estresse (SECPT-modificado) exatamente 25 min antes do início do teste final de memória com os 40 pares de palavras, enquanto a outra metade dos participantes foi submetida a uma condição controle. Cortisol salivar e respostas em questionários foram usadas para aferir a eficácia da indução de estresse. Os resultados mostram que o estresse foi induzido com sucesso no grupo experimental. Os participantes lembraram mais pares de palavras que haviam sido estudados por prática de lembrar do que por releitura, replicando achados da literatura. Inesperadamente, o efeito de prática de lembrar foi maior para pares de palavras fáceis do que difíceis. Além disso, a maioria dos participantes indicou maior expectativa de recordação para itens estudados por prática de lembrar do que por releitura, condizente com a performance real, mas na direção oposta aos resultados presentes na literatura, que indicam uma má calibragem metacognitiva em amostras similares. O efeito protetivo da prática de lembrar, como reportado em Smith et al. (2016), não foi replicado, possivelmente por cause de um efeito chão no condição de releitura. Os resultados, contudo, relevam uma interação não esperada na condição de prática de lembrar, a de que o estresse melhorou a performance de itens difíceis, mas piorou a de itens fáceis. Fornecemos uma descrição provisória desse achado, sugerindo que a prática de lembrar aplicada a itens difíceis adiciona um atributo de estresse ao traço de memória correspondente, cuja evocação é posteriormente facilitada sob estresse (análise condicionalizada). Os resultados sugerem que o impacto do estresse na memória depende tanto da estratégia de ensino adotada quanto da dificuldade intrínseca do item a ser aprendido.

Palavras-chave: prática de lembrar; recordação com pistas; estresse; cortisol; dificuldade do item.

Resumo Expandido

Pesquisas sobre memória humana mostram que a prática de lembrar –que é uma estratégia de estudo que consiste em utilizar o tempo de reestudo para evocar memórias já armazenadas– melhora a memória de longo prazo, e que a indução de estresse antes da evocação prejudica a performance da memória, por alterar o funcionamento do hipocampo, devido à saturação de receptores glicocorticoides pelo cortisol, hormônio liberado em consequência do estressor. Um estudo recente de Smith et al. (2016) mostrou que esse efeito deletério do estresse não acontece quando o conteúdo havia sido codificado via prática de lembrar. Neste estudo, nós buscamos replicar esse efeito protetivo da prática de lembrar sob estresse, controlando a dificuldade do item. Os participantes (59 alunos de graduação) estudaram 40 pares de palavras Suaíli-Português. Metade dos pares foram, então, estudados por releitura, enquanto a outra metade foi estudada por prática de lembrar. Metade dos pares em cada condição tinha memorabilidade alta (fáceis), enquanto a outra metade tinha memorabilidade baixa (difíceis), segundo dados extraídos de um recente estudo de normas suaíli-português. Os participantes foram questionados, ao final da sessão, sobre qual estratégia de estudo (releitura vs. prática de lembrar) traria melhores resultados em um teste subsequente, com o interesse em verificar o comportamento metacognitivo dos participantes sobre estratégias de estudo. Uma semana depois, os participantes retornaram para o teste final de memória dos 40 pares estudados. Metade dos participantes foi submetida a um protocolo de indução de estresse denominado SECPT-modificado. O protocolo envolvia tanto elementos de avaliação social (filmagem do participante) quanto de indução fisiológica de estresse (imersão de uma das mãos em água gelada, entre 0 e 2 ºC, por até 2 min). A indução do estresse ocorreu 25 min antes do início do teste final de memória com os 40 pares de palavras, momento em que acontece o pico do cortisol no organismo. A outra metade dos participantes foi submetida a uma condição controle, na qual a água estava na temperatura ambiente e não havia elementos de avaliação

social envolvidos na tarefa. O cortisol salivar e respostas em questionários de autorrelato de estresse e ansiedade foram usados para aferir a eficácia da indução de estresse. Os resultados mostram que o estresse foi induzido com sucesso no grupo experimental, mas o efeito principal do estresse na memória não foi achado entre grupos. Os participantes lembraram mais pares de palavras que haviam sido estudados por prática de lembrar do que por releitura, replicando achados da literatura. Inesperadamente: (1) o efeito de prática de lembrar foi maior para pares de palavras fáceis do que difíceis, o oposto ao hipotetizado, nos levando a sugerir que dificuldade da tarefa e dificuldade do item tem efeitos distintos na codificação e evocação da memória, e (2) a maioria dos participantes indicou expectativa de que lembraria mais itens estudados por prática de lembrar do que por releitura, condizente com a performance real, mas o oposto de resultados presentes na literatura, que indicam uma má calibragem metacognitiva em amostras similares. O efeito protetivo da prática de lembrar, como reportado em Smith et al. (2016), não foi replicado, possivelmente por causa de um efeito chão no condição de releitura e o não efeito significativo do estresse entre grupos. Os resultados, contudo, relevam uma interação não esperada na condição de prática de lembrar, a de que o estresse melhorou a performance de itens difíceis, mas piorou a de itens fáceis. Fornecemos uma descrição provisória desse achado, sugerindo que a interação entre itens difíceis e a prática de lembrar adicionou um atributo relacionado ao estresse ao traço de memória do item, de forma que em situação de estresse em teste posterior, a evocação desse traço seria facilitada (análise condicionada). Os resultados sugerem que o impacto do estresse na memória depende tanto da estratégia de ensino adotada quanto da dificuldade intrínseca do item a ser aprendido.

Palavras-chave: prática de lembrar; recordação com pistas; estresse; cortisol; dificuldade do item.

Introduction

Human memory is more than a mental box in which we keep records of our past; it is a complex system that recruits different areas of the brain, thus activating many cognitive and neural networks, representing one of the most remarkable abilities of human behavior. Memory is influenced by several different biological, psychological and environmental factors (Baddeley, Eysenck, & Anderson, 2015). Kandel (2001) thoroughly describes the molecular biology of memory and states that long-term memory (LTM) involves a specialized brain system in the medial temporal lobe, the hippocampus, and also the activation of gene expression, protein synthesis, neuronal growth and consequent formation of new connections. In fact, research on human memory has been conducted for over a century (Gates, 1917; Müller & Pilzecker, 1900; Spitzer, 1939) and still presents many open questions (Jaeger, Galera, Stein, & Lopes, 2016; Roediger, 2008). In educational settings, memory research is of great relevance, as it enables the assessment of learning strategies currently employed by educators and students (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Memory research in educational settings also allows the development of methodologies and materials that can help students encode and retrieve course-relevant information more effectively.

Retrieval practice

Among all different kinds of learning strategies, the most popular involves reading the to-be-learned materials several times (Karpicke, Butler, & Roediger, 2009). However, retrieving information from memory, rather than rereading the same information, is both a potent memory booster and also a learning event in itself (Bjork, 1994). People's beliefs about learning, usually built on faulty assumptions about memory functioning, often leads to the use of suboptimal learning strategies (Bjork, Dunlosky, & Kornell, 2013). Students' poor metacognitive awareness may be due to perceptual fluency (Bjork, 1999), which represents the sense of familiarity that increases when they read several times the same information, giving students a false perception of competence. This metacognitive illusion has great impact on the decision of which strategy is to be used, and students usually choose to read several times the to-be-learned information, although this strategy is considerably less effective than the retrieval practice (Roediger & Karpicke, 2006b). Educators also tend to be fonder of using learning strategies that encourage elaborative studying, instead of applying techniques that require students to practice reconstructing knowledge (Karpicke & Blunt, 2011). This is probably so because retrieving information on a test is usually considered a neutral event that measures the amount of information learned over a specific amount of time but does not stimulate or produce learning, and because both teachers and students tend to see tests as a bother. Therefore, tests are widely used in educational settings only as a means of longitudinally measuring students' cumulative knowledge.

However, recent systematic reviews (Adesope, Trevisan, & Sundararajan, 2017; Eisenkraemer, Jaeger, & Stein, 2013) and a large body of cross-disciplinary studies including applied linguistics and cognitive psychology (Barcroft, 2007; Barcroft, 2015; Bjork, 1975, 1988; Carpenter, Pashler, & Vul, 2006; Carrier & Pashler, 1992; Hogan & Kintsch, 1971; Karpicke & Roediger, 2008; McDaniel & Masson, 1985; Roediger & Karpicke, 2006a, 2006b; Spitzer, 1939; Tulving, 1967; Wheeler & Roediger, 1992) have shown that tests do much more then measuring learning; they can also enhance learning. As asserted by Storm, Bjork, and Storm (2010, p. 244): "Not only does information that has been tested become more recallable in the future than it would have been otherwise, that information, if retrieved, becomes more recallable than if such a test was replaced by an additional study opportunity".

Recent studies have repeatedly shown that long-term memory is affected positively by retrieval practice tests (Abel & Roediger, 2017; Roediger & Butler, 2011; Smith, Roediger, & Karpicke, 2013). Researchers have called this phenomenon the *testing effect* (Roediger &

Karpicke, 2006a; Wheeler & Roediger, 1992) or the *retrieval-practice effect* (Karpicke, Lehman, & Aue, 2014). Retrieval practice (RP), the study strategy that involves recalling information previously studied in order to aid future recall, is considerably more effective than other study strategies, such as study practice (SP), which involves rereading the to-be-learned materials, and concept mapping, which involves establishing elaborate relationships among the to-be-learned materials (Karpicke & Blunt, 2011). Retrieval-practice effects are observed even when retrieval practice occurs under divided attention (Buchin & Mulligan, 2019) and even when the final recall test occurs under conditions of stress (Smith et al., 2016; Szőllősi et al., 2017). The RP effect is modulated by retrieval difficulty during the encoding phase. Pyc and Rawson (2009) showed that recall rates in a delayed test are higher when the interval between retrieval episodes is long (difficult retrieval condition) compared to when the interval is short (easy retrieval condition). Studies show that RP effects are modulated by the retrieval difficulty during encoding, as participants recall rates are higher in a delayed test when the learning task was more difficult. The RP effect is also modulated by individual differences in participants characteristics (fluid intelligence) and stimulus characteristics (item difficulty) (Minear, Coane, Boland, Cooney, & Albat, 2018).

Karpicke and Roediger (2008) shows that RP produces large positive effects on long-term memory storage and retrieval, provided that the testing task promotes the opportunity to try to recall and reconstruct the information previously studied. That is because not any kind of testing produces learning, as not all test procedures include practicing retrieval, like when students take the test with prompt access to the relevant material (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008).

Experimental paradigms on the testing effect usually comprises three phases: (1) the study phase, in which participants are exposed to the to-be-learned materials (e.g., cue-target word pairs, such as *wingu-cloud*); (2) the practice phase, in which the to-be-learned materials undergo either restudy (e.g., re-reading *wingu-cloud*) or retrieval practice (e.g., recalling *cloud* given *wingu*-_____ or *wingu-cl*___); and (3) the criterial test phase, encompassing all stimuli presented on the study phase (e.g., recall *cloud* given *wingu*-_____ in a cued-recall paradigm).

Accounts of the retrieval-practice effect

Several hypothesis have been put forward to account for the retrieval-practice effect. In the following, we summarize the accounts most relevant to the present study.

Transfer-appropriate processing. Morris, Bransford, and Franks (1977) argued that the more similar the processing conditions in which the person encodes and retrieves information, the higher the probability of recall. Some findings support this account of the retrieval practice effect by showing that performance on the final test was better when tests formats during practice and final retrieval were identical (e.g., Johnson & Mayer, 2009). Although processing similarity may indeed modulate the testing effect, it may not be necessary to elicit the effect, as some studies have shown that retrieval practice elicits the highest improvement when participants were given a free-recall test during practice blocks, regardless of the similarity between intervening and final test formats (Carpenter & DeLosh, 2006; Glover, 1989).

Retrieval effort. Bjork (1975) compared the retrieval process with levels of cognitive processing (Craik & Lockhart, 1972), proposing that successful and effortful retrieval promotes a deeper level of processing and longer lasting memory retention than shallower and less difficult retrieval. The deeper the level of processing, the greater the strength given to the target. Response times (RT) are usually used as a measures of effort, with slower RTs indicating greater retrieval effort (Vaughn, Rawson, & Pyc, 2013). Bjork (1994) emphasized the need to introduce desirable difficulties in learning strategies. In a recent meta-analytic review, Rowland (2014) summarized evidence supporting the retrieval effort hypothesis.

Theory of disuse. In a subsequent development of the retrieval effort hypothesis, Bjork and Bjork (1992) posited two different types of strengths related to memory: (1) storage strength, that is relative to how well an item is learned and has no direct effects on memory performance; and (2) retrieval strength, that is directly involved in the probability of recalling the target given a related cue. New information in recent use has a high retrieval strength, although its storage strength is low, because of the limited number of times the information was accessed. By contrast, the ZIP code of your last home address has a high storage strength for it was accessed and used for a long time in the past but has a low retrieval strength in the present because of its disuse. They stated that the act of retrieving information is a memory modifier that facilitates subsequent successful recall by incrementing both memory strengths. Another assumption is that retrieval might be modulated by the difficulty involved during encoding tasks: the more difficult the task, the greater the benefits in delayed retrieval. This claim has been largely supported by a recent meta-analysis (Rowland, 2014).

Distribution-based bifurcation model. Kornell, Bjork, and Garcia (2011) elaborated on the premises of the theory of disuse by assuming that retrieval practice bifurcates the distribution of to-be-learned contents into strong and weak items. Theoretically, items that are tested without feedback, during the encoding phase, may pertain to two different classes of memory strength, those that are highly boosted when successfully recalled and those that are not boosted at all when not recalled. Items that were restudied by rereading are all boosted, but not as strongly and below the strength of unsuccessfully recalled items during retrieval practice blocks. The suggested explanation states that tests not only enhance learning, but also prevents forgetting because strong memories last longer. According to the bifurcation model, the presence of feedback after a recall attempt at encoding would hinder the bifurcation, or at least reduce the test-delay interaction – the longer the delay, the higher the advantage of retrieved (strong trace) over non-retrieved items (weak trace) –, because when feedback is given after an unsuccessful recall attempt, as shown by Kornell, Hays, and Bjork (2009), more learning is produced compared to restudying the same information without a retrieval attempt. These items could then reach the recall threshold on the delayed memory test. One important implication of feedback after a recall attempt during encoding is that it produces the best performance on both immediate and delayed memory tests.

Elaborative retrieval. Carpenter (2009) proposed that information activated during the attempt to retrieve a target may activate an elaborative semantic network of related concepts, creating multiple cue-driven pathways that lead to the correct target, aiding later retrieval. This hypothesis was based on spreading activation theories of memory (Collins & Loftus, 1975) . The elaborative retrieval hypothesis states that memory strength and semantic elaborations are modulated by the difficulty of the task. When the tasks are more difficult, target information is not promptly available, forcing a more extensive memory search and, thus, generating more semantic elaborations linked to the target. On the other hand, easier tasks during retrieval practice, in which the target is more evident (e.g., when the target is partially presented as a cue, or when the cue is highly associated to the target), hinders the generation of semantic associated information and, consequently, does not aid future recall.

Episodic context account. Karpicke et al. (2014) proposed an altogether different account from the ones reviewed above. Building on an influential temporal context model (Howard & Kahana, 2002), Karpicke et al. (2014) assumed that memory representations include both item and context information. Context may be external (e.g., settings) or internal (e.g., mental state). According to the episodic context account, context features originally stored with the target item are reinstated whenever retrieval is attempted, and the original memory trace is updated with this new context information. In this account, the enhancement in performance obtained by retrieval practice is attributed to the multiple temporal context representations encoded with the target, considering that retrieval itself drives context change, and that every time a target is retrieved, context information associated to the target is altered, incorporating new context elements, and thus creating unique composite features linked to the target. This updated, composite representation would allow constraining the mental search set for the target. The search set is the group of potentially recallable items that match the features of a given cue. Such restricted search set, afforded by repeated retrieval attempts, facilitates the match between cue and target, and consequently improves performance on subsequent memory tests. Evidence supporting the episodic account comes from studies showing that memory performance is higher when materials are studied in spaced schedules (Kang, 2016), creating new context-related elements to the memory trace, from research about list interference in free recall paradigms (Jang & Huber, 2008), showing the impact of context change in memory performance, or from studies in which cue difficulty is manipulated during practice (Carpenter & DeLosh, 2006), showing that the retrieval practice effect is less strong with easier cues than it is with more difficult (fewer) cues, in which context reinstatement would be necessary for successful performance.

In the aforementioned accounts, retrieval practice is thought to boost subsequent retrieval as a result of (a) task similarity, (b) deeper encoding, (c) activation of elaborate semantic cues, (d) restriction of the search set or any combination thereof. Recent studies have begun to investigate whether retrieval practice can counteract, by any of these mechanisms, the memory-impairing effects of concurrent tasks (Buchin & Mulligan, 2019) and psychosocial stress (Smith et al., 2016; Smith, Race, Davis, & Thomas, 2018; Smith & Thomas, 2018; Szőllősi et al., 2017).

Stress and retrieval

In his highly influential book, Selye (1956) fully describes his understanding of the *syndrome of just being sick*, which would later be defined by the word *stress*, meaning the general response of the body to any demand for change. The physiological stress response aims at maintaining homeostasis, a concept defined by Cannon (1929) as being an adaptational effort of the physiological systems to maintain an acceptable level of functioning when facing “fight or flight” conditions. Following the Selyean tradition, the term stress will be used here to refer to a physiological response caused by a stressor (a stimulus that triggers the cascade of stress-related events in the body). According to Girdano, Dusek, and Everly Jr (2012), there are two types of stressors: (1) biogenic stressor, which possesses inherent physiological properties that enables it to directly initiate a stress response, bypassing the neocortex higher interpretative mechanisms (e.g., drinking coffee, holding very cold items, exercising), and (2) psychosocial stressor, which depends on higher neocortical processing and involves cognitive appraisal and affective integration of the stressor itself (e.g., public speaking, traffic jams). Not all psychosocial stimuli turn into a psychosocial stressor, because this change depends on individual perception and interpretation of these stimuli: “like beauty, the stressor resides in the eye of the beholder” (Everly Jr & Lating, 2019, p. 241). As Hans Selye once noted, “it is not what happens to you that matters, but how you take it”.

Potentially stressful events are common in everyone’s life. Due to its relevance to people in general, stress and its main hormone, cortisol, are widely investigated topics (Gagnon & Wagner, 2016; Lonergan, Olivera-Figueroa, Pitman, & Brunet, 2013; Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012; Shields, Sazma, McCullough, & Yonelinas, 2017; Vogel & Schwabe, 2016; Wolf, 2017). Stress neuroendocrine responses activates both the sympathetic nervous system (SNS) and the hypothalamic–pituitary–adrenal (HPA) axis, respectively leading to (1) the release of noradrenaline and adrenaline (NA) by the adrenal medulla and (2) the secretion of glucocorticoids (cortisol) by the adrenal cortex, both into the bloodstream (Charmandari, Tsigos, & Chrousos, 2005; Wolf, 2017). The hippocampus is a medial temporal lobe structure that contains both types of corticosteroid receptors, mineralocorticoid receptors (MRs) and glucocorticoid receptors (GRs), and that plays a critical role in binding information into coherent declarative memory representations and is crucial for the retrieval of such memories (Kim, Song, & Kosten, 2006; Yassa & Stark, 2011).

Stress-induced biochemical cascades and cortisol release can have a direct impact on the hippocampus and the amygdala, brain structures pertaining to the limbic system, which are highly involved in memory and emotional processing (Kim & Diamond, 2002; Roozendaal & McGaugh, 2011). Cortisol is a steroid hormone that can cross the blood-brain barrier and bind to GRs in the hippocampus. Excessive amounts of circulating cortisol can saturate these receptors, altering hippocampal function, and consequently, interfering with retrieval-related processing in this region, thus modulating encoding and retrieval of long-term memories (Schwabe et al., 2012).

A considerable number of articles assessed the relationship between cortisol and different stages of memory processing (encoding, storage and retrieval). Empirical evidence, however, has been mixed, with studies showing that episodic memory can be enhanced (Hupbach & Fieman, 2012; Zoladz et al., 2014), impaired (Boehringer, Schwabe, & Schachinger, 2010; Gagnon, Waskom, Brown, & Wagner, 2019; Goldfarb, Mendelevich, & Phelps, 2017; Schonfeld, Ackermann, & Schwabe, 2014; Schwabe & Wolf, 2014), or unaffected (Hidalgo et al., 2015; Pulopulos et al., 2013; Schoofs & Wolf, 2009) after stressful events. A recent review has shown that memory performance of old people is less influenced by stress because of age-related decrease of GRs density and sensitivity in the hippocampus and that women’s susceptibility to stress is highly dependent of menstrual hormones (Hidalgo, Pulopulos, & Salvador, 2019). In addition to participants’ sex and age, timing is also crucial for understanding how stress modulates memory (Schwabe & Wolf, 2014), because neuroendocrine stress response plays different roles depending on the memory process affected. The fast stress response occurs seconds after the stressor and involves the release of (nor)adrenaline, which increases alertness and facilitates memory encoding. The slow stress response, on the other hand, occurs several minutes after the stressor and involves the release of cortisol, which impairs the retrieval of consolidated memories (Schwabe et al., 2012). According to Shields et al. (2017), when the stressor occurs at encoding, the event will be better stored or consolidated, usually involving some level of aversive conditioning, as when a person goes through a robbery and is able to remember the event in details for a lifetime. However, when the stressing event takes place just before the attempt to retrieve previously consolidated information from memory, retrieval is impaired.

Theories about the impact of stress in memory are generally related to the consolidation hypothesis proposed by Müller and Pilzecker (1900), which states that recent encoded items persist in a fragile state and that consolidation happens over time. Thus, newly learned information is susceptible to exogenous modulatory treatments (e.g., learning other new items, emotional arousal). McGaugh (2000) say that stress may play an important role in regulating the consolidation of the information learned, because the interactions among different neural systems involved in the stress response are beneficial to the consolidation process and help stabilize the information in long-term memory. Stress should enhance memory encoding because recently learned information benefits from both the memory formation mode set after the stressor, through rapid non-genomic MR-mediated activity, and from the memory storage mode, elicited by the slow GR-dependent genomic effects of cortisol in the brain (Vogel & Schwabe, 2016). Schwabe, Nader, and Pruessner (2014) propose that when the memory is reactivated, it becomes labile and once again susceptible to modification. Stress, in this case, would enhance future performance by adding emotional arousing elements to the memory trace during the reconsolidation process. Considering that both memory formation and storage modes inhibit retrieval, stress induced before a recall attempt should impair performance.

Retrieval practice and stress

A recent study has suggested that retrieval practice is successful at creating strong memory representations that are resilient to stress (Smith et al., 2016). In their study, Smith et al. (2016) manipulated learning strategy (RP vs. SP) and stress condition (stressed vs. control) in a between-subject design. Stimuli consisted of 30 nouns and 30 nameable images. Participants were shown the stimuli and were asked to type in a sentence that included each item. Then they either restudied the items or freely recalled them (without feedback). Participants returned 24 hr later for the final free recall test. Stress was induced at the beginning of the session with the widely used Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993). The final test took place 25 min after stress induction. The results showed that participants recalled more RP items than SP items. More importantly, stress impaired performance only for SP items; recall of RP items was not affected by stress. The finding that retrieval practice appears to reduce the memory-impairment effects of stress may play an important role in applied settings. Retrieval practice can be used, for example, to reduce student’s blank-outs during high-stakes tests.

Those results, however, have not been easy to replicate. Szőllősi et al. (2017) tried to extend Smith et al. (2016) findings by using a within-subject design for the learning strategy manipulation and a more realistic 7-day delay between the study and criterial test phases. Although Szőllősi et al. (2017) did find the retrieval practice effect, they did not find the effect of stress on memory, nor did they find the predicted strategy by stress interaction. Smith, Davis, and Thomas (2018) also failed to replicate Smith et al. (2016). Smith et al.’s (2018) results show that retrieval practice yielded better recall rates than study practice, but performance was equally impaired by stress, regardless of learning strategy. Both Szőllősi et al. (2017) and Smith, Davis, et al. (2018) used Swahili-English word pairs in a cued-recall paradigm, as opposed to lists of nouns in a free recall paradigm in Smith et al. (2016) original study. As argued by Smith, Davis, et al. (2018), foreign vocabulary learning requires more complex cognitive processes and retrieval practice is better for stimuli of low complexity, a limitation only evident under stress.

Current study

The aims of this study are (1) to replicate and extend Smith et al. (2016)'s findings to confirm that retrieval practice creates memories resistant to stress; (2) to evaluate the impact of item-difficulty on the RP effect by manipulating the to-be-learned word pairs difficulty within-subjects, considering that retrieval effort plays an important role in most accounts of the retrieval practice effect and that retrieval practice should benefit harder targets more than easier targets; and (3) to evaluate the metacognitive awareness of students about RP.

Based on the reviewed literature, we hypothesize that (1) recall will be significantly higher for retrieval practice (RP) targets relative to study practice (SP) targets, (2) the deleterious effects of stress on memory will be found for SP targets but not for RP targets, (3) that harder targets will be recalled at a higher rate than easier ones for RP condition, and (4) that most participants will choose SP as being more effective than RP.

Method

Participants

Sixty-three healthy young men were recruited among University of Brasília (UnB) undergraduates to voluntarily participate in the experiment. Three participants were excluded from the study sample because cortisol analyses could not be performed with their provided saliva sample. Further, one participant was excluded because he was not following the experimental instructions. Therefore, we report results of 59 male participants (age range 19-37, M_{age} = 22.5, *SD* = 3,67). Participants were randomly assigned to one of the two groups: Control (*N* = 29, M_{age} = 22.6, *SD* = 3,76) or Stressed group (*N* = 30, M_{age} = 22.4, *SD* = 3,65). Before starting the experiment, participants read and signed the consent form (Appendices D and E). This research was approved by the Ethical Committee of the Institute of Social and Human Sciences of University of Brasília (Appendix C). Women were not included to avoid the confounding influence of the period of the menstrual cycle on cortisol-related memory performance (e.g., Shields et al., 2017). Participants were all undergraduate students from the University of Brasília and native Portuguese speakers. Individuals who had previous knowledge of the Swahili language, had a history of neurologic or psychiatric disorders, were taking psychotropic medication, or were smokers could not be participants. In the beginning of each session, participants answered a questionnaire that assessed whether they had drunk coffee or any other energy drink (6 hrs before the beginning of the session), smoked cigarettes (12 hrs), had a meal (3 hrs), smoked cannabis (24 hrs), taken illegal synthetized drugs (7 days), drunk alcoholic beverages (12 hrs), taken any controlled legal drugs (12 hrs), been previously diagnosed with chronic stress, or been exposed to the Swahili language. No participants met any of these exclusion criteria.

Experimental Design

The study followed a 2 (encoding strategy) × 2 (target difficulty) × 2 (groups) factorial design. Encoding strategy (hereafter, *strategy*) was manipulated within participants and refers to the study technique applied in distinct study blocks: Study Practice (SP) or Retrieval Practice (RP). Target difficulty (hereafter, *difficulty*) was also manipulated within participants and refers to the pre-experimental memorability of the study items, classified either as Easier (higher memorability) or harder (lower memorability) items. Group was manipulated between participants and refers to the stress manipulation: in the control group, participants followed a non-stressful protocol, whereas in the stressed group, participants followed a stressful protocol (more details below).

Materials and tasks

Stimuli. Participants were exposed to 40 Swahili-Portuguese word pairs (Appendix A) selected from Brazilian Portuguese norms by Lima and Buratto (2019), adapted from the original Swahili-English norms (Nelson & Dunlosky, 1994). Stimuli were split into two lists based on memorability scores, resulting in 20 items for the RP condition and 20 items for the SP condition. Half of the items in each list were easier (with high memorability scores, M = .60, SD = .10); the other half were harder (with low memorability scores, M = .24, SD = .05). The assignment of word pairs to encoding strategy condition (RP vs. SP) was counterbalanced across participants.

Stress induction protocol (modified SECPT). Stress was induced during the final test session with a modified version of the Socially-Evaluated Cold Pressor Test (SECPT). The SECPT is a simple, effective, and widely used procedure that reliably activates the HPA axis, resulting in significant increases in cortisol levels measured by saliva samples (Schwabe, Haddad, & Schachinger, 2008; Schwabe & Schachinger, 2018). In our modified SECPT, participants were told to keep their non-dominant hand and wrist submerged in a recipient filled with ice water (0–2 ºC) from 1 to 2 minutes. Prolonged cold water stimulation activates both the SNS, increasing blood pressure and heart rate, and the HPA, increasing cortisol levels (Schwabe, Bohringer, Chatterjee, & Schachinger, 2008). In addition to the physiological stress induced by cold water, the SECPT procedure also includes a psychosocial stress component. During cold water hand immersion, participants were told to look directly to a camera that would record their facial reactions for subsequent evaluation by an external committee. The experimenter, a female testing male participants, what is shown to elicit higher psychological stress responses (Duchesne, Tessera, Dedovic, Engert, & Pruessner, 2012), took notes during the protocol while monitoring participants' compliance to the instructions. After the hand submersion phase, participants were instructed to watch a 20-min excerpt from the TV series *The Office*. Halfway through the episode, participants were told by surprise to carry out a mental subtraction task. Specifically, they were told to begin with 100, subtract 7, and say aloud the result, then subtract 7 again, and say aloud the result, (e.g., 93, then 86), repeating the procedure for 1 min or until they reached 0. This latter mental arithmetic task is not part of the original SECPT procedure; instead, it is part of another commonly used stress-induction protocol known as the Trier Social Stress Test (Allen et al., 2017; Kirschbaum et al., 1993). The mental arithmetic task adds yet another social-evaluative and unpredictability factors to the SECPT protocol. A similar modified SECPT procedure has been shown to strongly activate the HPA axis (Boyle et al., 2016). In the control version of our modified SECPT procedure, applied in the control group, participants completed similar steps, except that the water was warm (23–25ºC) and no social evaluative elements were used (i.e., no video recording and no surprise mental arithmetic task). The experimenter, also a woman, did not take notes during the protocol, remaining in the room just to check participants' compliance to instructions.

Physiological measurements. Blood pressure (systolic and diastolic) and heart rate were measured with an automatic blood pressure monitor (HEM-7130; Omron Healthcare Brazil, São Paulo). This device was chosen because it is compact, fully automated and has been validated against measurements from standard mercury sphygmomanometers (Takahashi, Yoshika, & Yokoi, 2015). Salivary cortisol samples were collected with a chewable synthetic swab (Sarstedt Cortisol Salivette® code blue; Nümbrecht, Germany). According to manufacturer specifications, Cortisol Salivette® yields a high sample recovery rate upon centrifugation, achieving high recovery rates even with small samples (e.g., 200 μL) (www.sarstedt.com/en/products/diagnostic/salivasputum/product/51.1534.500/, accessed in August 2019). Also, according to the manufacturer, tests with adults revealed an average saliva recovery volume of 1.1 +/− 0.3 ml. In addition, the efficacy of Cortisol Salivette® was established in the volume of cortisol recovering, showing close to 100% recovery rates, regardless of the saliva volume or test procedure. In fact, recent studies indicate that Cortisol Salivette® can yield acceptable recoveries (Buttler et al., 2018; Groschl, Kohler, Topf, Rupprecht, & Rauh, 2008).

Self-report stress questionnaire (BAI). Subjective stress was measured with the Beck Anxiety Inventory (BAI), a 21-question, multiple-choice, self-report inventory used for measuring anxiety. BAI is widely used in research and has been validated in its Brazilian Portuguese version (Cunha, 2001). Participants were instructed to answer BAI, before the criterial memory test, considering only the time scope of the current session. This was done to assess their self-reported stress only about the SECPT manipulation between groups.

Procedure

Participants attended two individual sessions (Figure 1). In the beginning of each session, several baseline physiological measurements were taken, namely, heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), and salivary cortisol (SC). Participants also completed the self-report stress questionnaire (BAI). Both sessions took place in the same room, which was equipped with air conditioner, the cold pressor apparatus and a computer. The room is part of the Laboratory for Psychological Research in Humans (LIPSI), at the Institute of Psychology of the University of Brasília.

Session 1, referred to as the Encoding Phase in this study, contained the memory encoding task and session 2, referred to as the Criterial Test Phase in this study, occurred after a 7-day delay and contained the experimental intervention (SECPT), as well as the cued-recall criterial memory test.

	Encoding phase (session 1) ………………..	7_{days} …..	Criterial test phase (session 2)
T_0	Read and sign consent form	T_8	Collect salivary cortisol sample (SC2)
T_1	Collect salivary cortisol sample (SC1)	T_9	Collect physiological measures 2
T_2	Collect physiological measures 1	T_{10}	Answer exclusion criteria questionnaire
T_3	Answer exclusion criteria questionnaire	T_{11}	Answer Beck anxiety inventory (BAI2)
T_4	Answer Beck anxiety inventory (BAI1)	T_{12}	SECPT (control or stressed)
T_5	Study all word pairs (40 WP)	T_{13}	Collect salivary cortisol sample (SC3)
T_6	Practice blocks (total of 6 RP and 6 SP blocks of 3 min each)	T_{14}	Collect physiological measures 3
	[A] RP blocks (word pairs 1-20)	T_{15}	Answer Beck anxiety inventory (BAI3)
	[B] SP blocks (word pairs 21-40)	T_{16}	Criterial cued-recall test (40 WP)
	Block sequence:	T_{17}	Sign the debriefing document
	either [A], [B], [A], [B], [A], [B]…		
	or [B], [A], [B], [A], [B], [A]…		
T_7	Sign the attendance document		

Figure 1. Experiment timeline.

In session 1 (Encoding phase), participants first read and signed the consent form, had their baseline physiological measures collected (HR1, SBP1, DBP1 and SC1), and filled out the BAI questionnaire (BAI1). Next, participants studied all 40 word pairs, presented randomly in the computer screen for 6 s each. They were instructed to read the word pairs with attention. Then, participants completed 6 study practice (SP) blocks and 6 retrieval practice (RP) blocks, in alternating order (Figure 1, T6), each block comprising 20 word pairs (Appendix A, tables A1 and A2). Primacy and recency effects were controlled by randomizing the order of stimuli within lists and across participants. In both SP and RP trials, participants were instructed to type in the Portuguese translation as soon as they saw the Swahili word. In SP trials, both Swahili and Portuguese words were simultaneously presented for 6 s on the screen, and participants had up to 4.5 s to type in the Portuguese words. In RP trials, participants had 4.5 s to recall and type in the translation of the Swahili word presented on the screen as a cue. The target, Portuguese word, was only presented for the last 1.5 s and served as feedback. Participants completed 30-s distractor tasks between blocks. The distractor task consisted of simple multiplications and was used mainly to prevent participants from mentally rehearsing the word pairs from the last block. After completing the 12 blocks, participants booked the next session, always between 12pm and 5pm, signed the checklist for the session (Appendix F) and were dismissed.

In session 2 (Criterial test phase), participants returned to the same room 7 days after session 1. Again, participants first had their baseline physiological measures collected (HR2, SBP2, DBP2 and SC2) and then filled out the BAI questionnaire with the instruction to consider the last 7 days when answering the questions (BAI2). Next, participants underwent the stress protocol (modified SECPT) or the control protocol. After completion of the stress protocol, participants filled out the BAI questionnaire for the third and last time, with the instruction to consider only the current session when answering the questions (BAI3). The delay between the stressor onset (T12) and the beginning of the criterial test (T16) was 25 min. This delay was motivated by research showing that 25 min is the average time for cortisol concentration to reach its peak in humans (Joels & Baram, 2009; Joels, Fernandez, & Roozendaal, 2011; Schonfeld et al., 2014; Schwabe et al., 2012; Schwabe & Wolf, 2014). Participants started the criterial test as soon as the physiological measures were collected, and salivary cortisol was sampled. For the criterial memory test, cues were presented one at a time, for 15 s each, on the computer screen, and participants were instructed to type in targets as soon as they had recalled them. Stimulus presentation and data collection were implemented in PsychoPy (Peirce et al., 2019). After the criterial test, participants read and signed the debriefing form (Appendix G).

Data Analyses

An alpha level of .05 was set for all statistical tests. When the assumption of sphericity was violated, as indicated by Mauchly's test, the Greenhouse–Geisser correction was applied to adjust for degrees of freedom (Greenhouse & Geisser, 1959). Measures of effect size were reported as Cohen's *d* (*t*-tests), partial eta-squared (η_p^2; ANOVAs), or log odds ratio (β; multilevel logistic regressions), when appropriate. Error bars in figures represent the 95% within-participant confidence intervals (Cousineau, 2005). Recall data for both sessions of the experiment were first obtained using formulas in Microsoft Excel (i.e., first-pass check on whether the participant typed the expected target word). The results from the Excel formulas were then compared to that of a completely blind judge who was trained to assess participants' responses on practice blocks and on the final criterial test. The judge was blind to participants' names, strategy used, item difficulty and groups. The agreement between the Excel formula and the blind judge was assessed using Cohen's kappa (κ) coefficient. The level of inter-rater agreement between judges was considered perfect, $\kappa = 1$, $p < .001$.

Results

Encoding phase (Session 1)

Physiological measures and self-reported stress. Independent-samples *t*-tests were conducted to compare data between groups (stressed vs. control). No differences in salivary cortisol concentration (SC1), systolic blood pressure (SBP1), diastolic blood pressure (DBP1), and heart rate (HR1) were found between groups, *t*s < 1.07, *p*s > .29. There was also no difference in self-reported stress scores (BAI1) between groups, $t < 1$, $p = .82$.

Target recall on practice blocks. Figure 2 depicts target recall across 5 practice blocks.[1] Although there were a total of 6 blocks for the retrieval practice, the first block was used to check and, if necessary, reinforce the instruction to type in the answer. Consequently, only data from blocks 2 to 6 were available for analyses.

A 2 (Difficulty) × 5 (Block) × 2 (Group) repeated-measures ANOVA revealed significant main effects of Difficulty, $F(1, 57) = 184$, $p < .001$, $\eta_p^2 = .76$, and Block, $F(4, 228) = 159$, $p < .001$, $\eta_p^2 = .74$, as well as a significant Difficulty × Block interaction, $F(4, 228) = 8.03$, $p < .001$, $\eta_p^2 = .12$. Main effect of Group and all other interactions were not significant, all *p*s > .37.

Paired-samples *t*-tests showed that participants recalled more targets on each subsequent practice blocks (Bonferroni corrected *p*s < .001), and that easier targets were successfully recalled more often than harder targets on each block, (Bonferroni corrected *p*s < .001). Paired-samples *t*-tests also revealed that the learning advantage for easier over harder targets increased throughout blocks: block 2, $M_{dif} = .17$, $t(58) = 7.96$, $p < .001$, $d = 1.04$, block 3, $M_{dif} = .22$, $t(58) = 8.20$, $p < .001$, $d = 1.07$, block 4, $M_{dif} = .27$, $t(58) = 10.5$, $p < .001$, $d =$

[1] Only RP data were included in recall and reaction time analyses for practice blocks. Analyses of SP data would be inappropriate here because SP targets were always on the screen while participants typed them in, thus reflecting cognitive and neuronal processes other than those related to memory retrieval.

1.37, block 5, $M_{\text{dif}} = .26$, $t(58) = 11.1$, $p < .001$, $d = 1.44$, block 6, $M_{\text{dif}} = .30$, $t(58) = 13.2$, $p < .001$, $d = 1.71$. This learning advantage for easier over harder targets, however, was similar for stress and control groups.

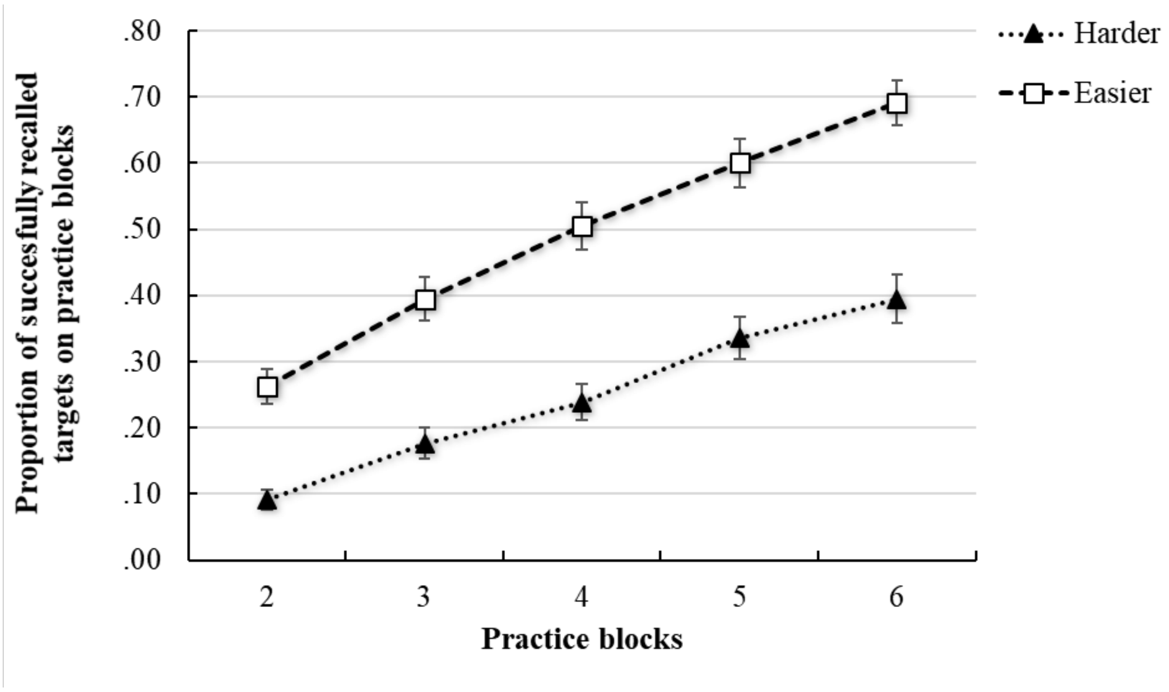

Figure 2. Target recall as a function of practice blocks and item difficulty.

RT (reaction time) on encoding phase. Figure 3 depicts mean reaction times across blocks 2 to 6 during the encoding phase. A 2 (Difficulty) × 5 (Block) × 2 (Group) repeated-measures ANOVA revealed significant main effects of Difficulty, $F(1, 24) = 22.7$, $p < .001$, $\eta_p^2 = .49$, and Block, $F(4, 96) = 15.8$, $p < .001$, $\eta_p^2 = .40$. Neither the main effect of Group nor any interaction terms were significant (all *p*s > .13). Paired-sample *t*-tests showed that, in every practice block, easier targets were recalled faster and by a larger number of participants than harder targets, block 2 ($n = 49$, $M = 1{,}974$ ms, $SD = 650$ ms vs. $n = 29$, $M = 2{,}250$ ms, $SD = 693$ ms), block 3 ($n = 54$, $M = 1{,}773$ ms, $SD = 455$ ms vs. $n = 40$, $M = 1{,}927$ ms, $SD = 539$ ms), block 4 ($n = 58$, $M = 1{,}611$ ms, $SD = 403$ ms vs. $n = 45$, $M = 1{,}844$ ms, $SD = 394$

ms), block 5 (n = 58, M = 1,603 ms, SD = 353 ms vs. n = 51, M = 1,835 ms, SD = 526 ms), and block 6 (n = 59, M = 1,535 ms, SD = 277 ms vs. n = 53, M = 1,729 ms, SD = 389 ms). Additional analyses on the automatization account (Racsmany, Szollosi, & Bencze, 2018) are provided in Appendix B.

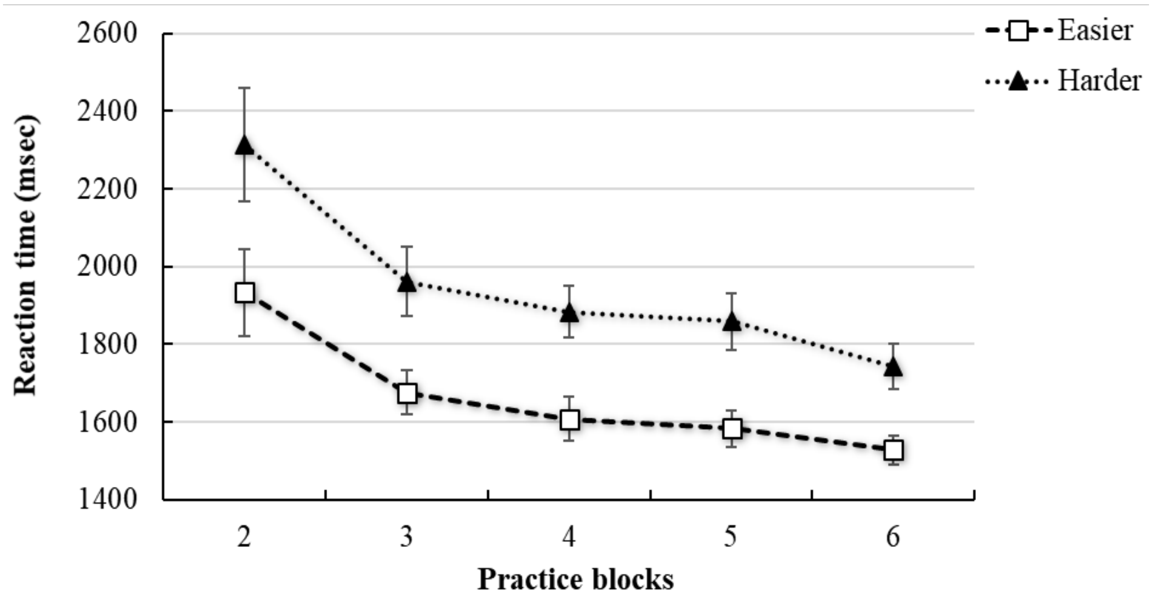

Figure 3. Reaction time for correct target recall as a function of retrieval practice blocks.

Judgement of learning (JOL). At the end of session 1, participants were asked to predict which one of the two learning strategies would be more effective for recall in a final memory test 7 days later. Most participants (62.7%) chose retrieval practice over study practice. A goodness-of-fit chi-square test showed that the choice of learning strategy was significantly different from chance, χ^2 (1) = 3.81, p = .05.

Criterial test phase (Session 2)

Physiological measures. Participants had their heart rate and blood pressure measured, as well as their saliva sampled both before starting the SECPT protocol and 25 min after its onset. Post-hoc *t*-tests were conducted to compare stress and control groups. Table 1 shows mean results for the physiological measures on the criterial test phase.

Cortisol responses. A 2 (Group) × 2 (Intervention: Pre- or Post-SECPT) mixed-factorial ANOVA on cortisol salivary readings showed significant main effects of intervention, $F(1, 57) = 47.5$, $p < .001$, $\eta_p^2 = .46$, Group, $F(1, 57) = 12.7$, $p = .001$, $\eta_p^2 = 0.18$, and Group × Intervention interaction, $F(1, 57) = 43.6$, $p < .001$, $\eta_p^2 = .43$. Post-hoc *t*-tests confirmed that cortisol concentration increased strongly in response to the SECPT in the stressed group, $t(29) = 7.72$, $p < .001$, $d = 1.41$, but not in the control group, $t < 1$, $p = .75$. Figure 4 illustrates these results. Following the SECPT procedure, the water temperature was significantly lower in the stressed group (M = 1.35 °C, SD = 0.60 °C) than in the control group (M = 24.2 °C, SD = 1.71 °C), $t(34.6) = 68.3$, $p < .001$. Accordingly, participants kept their hand underwater for a shorter period in the stressed group (M = 80.9 s, SD = 25.1 s) than in the control group (M = 106 s, SD = 21.1 s), $t(57) = 4.22$, $p < .001$.

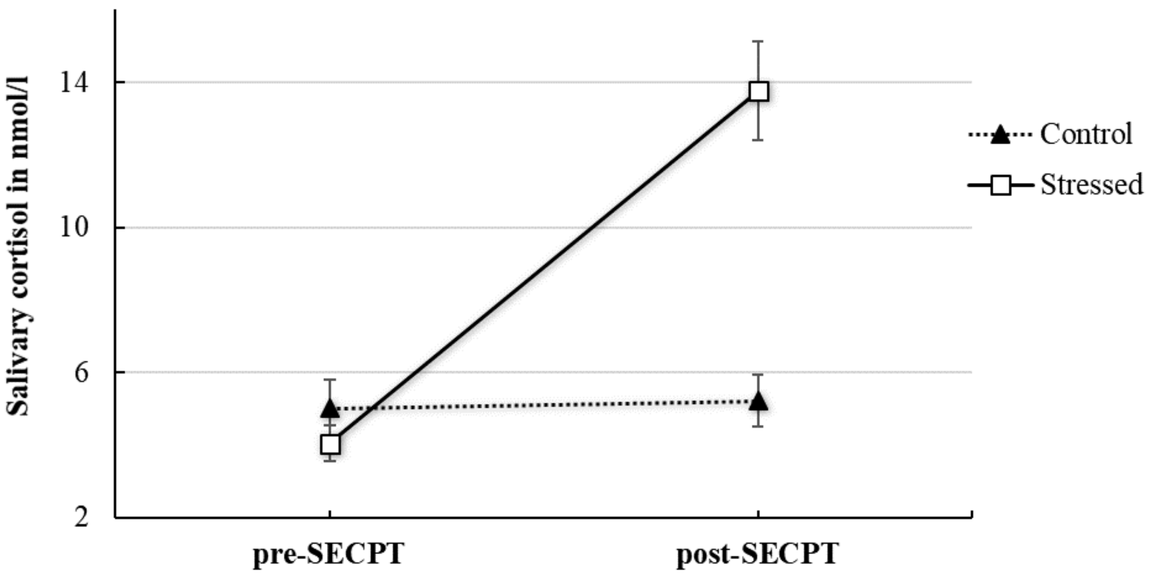

Figure 4. Cortisol concentrations at pre- and post-SECPT for control and stressed groups.

Blood pressure and heart rate. Considering pre-SECPT data, participants did not differ in systolic blood pressure (SBP2), diastolic blood pressure (DBP2) and heart rate (HR2; *t*s < 1, *p*s > .35). Post-SECPT data also showed no differences in systolic blood pressure (SBP3) and heart rate (HR3) *t*s < 1, *p* = .81. Groups differed, however, in post-SECPT diastolic blood pressure (DBP3), *t*(57) = –2.16, *p* = .04, with participants in the stressed group (*M* = 74.03 mmHg, *SEM* = 1.42) presenting higher diastolic blood pressure than participants in the control group (*M* = 70.21 mmHg, *SEM* = 1.03).

Self-reported stress (BAI). Post-hoc independent-samples *t*-tests showed no group differences in baseline (pre-SECPT, BAI2) scores (*t* < 1, *p* = .77). Post-SECPT scores (BAI3), however, differed between groups with higher scores in the stressed group (*M* = 6.57, *SEM* = 0.90) than in the control group (*M* = 2.76, *SEM* = 0.68), *t*(57) = 3.37, *p* = .001.

Table 1. Average physiological and self-reported stress measures.

	Control[e]		Stressed[e]	
Measures	Pre-SECPT	Post-SECPT	Pre-SECPT	Post-SECPT
Cortisol concentrations[a]	5.01 (0.81)	5.22 (0.72)	4.05 (0.49)	13.76 (1.37)
Systolic blood pressure[b]	119 (2.32)	113 (2.37)	115 (2.41)	113 (1.76)
Diastolic blood pressure[b]	73.3 (1.27)	70.2 (1.03)	73.7 (1.64)	74.0 (1.42)
Heart rate[c]	74.7 (2.43)	71.9 (1.98)	77.4 (2.27)	72.6 (1.96)
Self-reported stress (BAI3)[d]		2.67 (0.68)		6.57 (0.90)

[a] nmol/l. [b] mmHg. [c] bpm. [d] participants were instructed to respond the post-SECPT BAI questionnaire considering just the time of the session, to evaluate the difference in self-reported stress related to the experimental intervention (modified SECPT). [e] Standard errors of the mean are given in parentheses.

Target recall on criterial test. Criterial test data were submitted to a 2 (Strategy: SP or RP) × 2 (Difficulty: Easier or Harder) × 2 (Group: Control or Stressed) mixed-design

ANOVA. The ANOVA revealed two significant main effects: (1) Strategy, $F(1, 57) = 137$, $p < .001$, $\eta_p^2 = .71$, indicating that participants recalled more targets previously encoded via RP ($M = .36$, $SD = 0.20$) than targets encoded via SP ($M = .12$, $SD = 0.11$), and (2) Difficulty, $F(1, 57) = 147$, $p < .001$, $\eta_p^2 = .72$, indicating that easier targets were recalled more often than harder ones ($M = .34$, $SD = 0.19$ vs. $M = .14$, $SD = 0.11$). Figure 5 illustrates these results.

The main effect of Group and the Strategy × Group interaction were not significant, *p*s > .26. The three-way interaction, Strategy × Difficulty × Group, was significant, $F(1, 57) = 5.85$, $p = .02$, $\eta_p^2 = .09$. To better understand this three-way interaction, considering our interest in investigating the interaction of RP and stress, we analyzed the data separately for RP targets and SP targets by conducting two 2 (Difficulty) × 2 (Group) ANOVAs.

For RP targets, the ANOVA revealed a significant a main effect of Difficulty, $F(1, 57) = 115$, $p < .001$, $\eta_p^2 = .67$ (higher recall for easier targets). The Difficulty × Group interaction was also significant, $F(1, 57) = 10.8$, $p = .002$, $\eta_p^2 = .16$. Independent-samples *t*-tests showed that recall of easier targets was significantly impaired in the stressed group compared to the control group, $M_{\text{dif}} = –0.13$, $t(57) = –2.05$, $p = .045$. For harder targets, no group difference was found ($t < 1$, $p = .51$). The main effect of Group was not significant ($F(1,57) = 0.88$, $p > .35$). For SP targets, the ANOVA revealed a significant main effect of Difficulty, $F(1, 57) = 43$, $p < .001$, $\eta_p^2 = .43$ (higher recall for easier targets). Neither the main effect of Group, nor the Difficulty × Group interaction were significant (*p*s > .91).

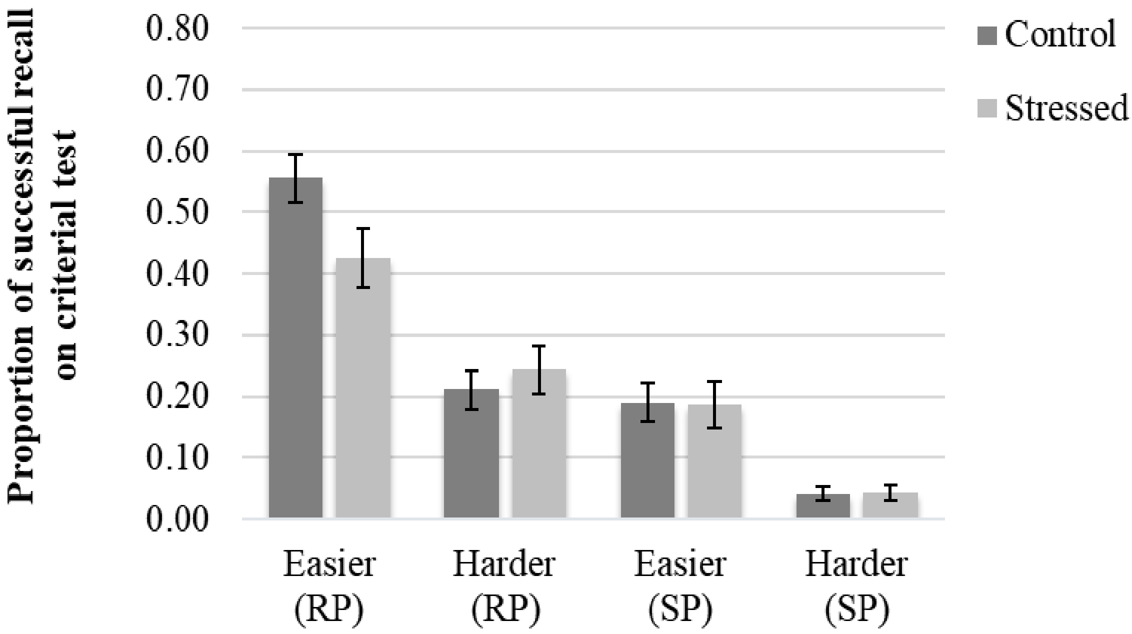

Figure 5. Cued-recall performance on the criterial memory test.

Reaction time (RT). Criterial test RT data were entered a 2 (Strategy) × 2 (Difficulty) × 2 (Group) mixed-design ANOVA. The ANOVA revealed a significant main effect of Strategy, $F(1, 17) = 13.3$, $p = .002$, $\eta_p^2 = .44$, showing that retrieval practice targets were recalled faster than study practice targets. All other main effects and interactions were not significant (all *p*s > .24). Considering that (1) the number of participants included in this 2 × 2 × 2 ANOVA was very low (control: $n = 10$; stressed: $n = 9$), and (2) reaction time is usually used as an auxiliary measure of recall difficulty (Vaughn et al., 2013), we analyzed data separately for Strategy and Difficulty factors in two 2 × 2 ANOVAs.

A 2 (Strategy) × 2 (Group) ANOVA, revealed a significant main effect of Strategy, $F(1, 42) = 14.6$, $p < .001$, $\eta_p^2 = .26$. Figure 6a depicts participants' reaction time for correct recall on criterial test, which was considerably shorter for retrieval practice targets ($M = 1{,}006$ ms, $SD = 974$ ms) than for study practice targets ($M = 2{,}176$ ms, $SD = 1{,}837$ ms), $t(44) = 3.85$, $p < .001$, $d = 0.58$). All other main effects and interactions were not significant (*p*s > .41).

A 2 (Difficulty) × 2 (Group) ANOVA, revealed a significant main effect of Difficulty, $F(1, 47) = 7.88$, $p = .007$, $\eta_p^2 = .14$. Figure 6b depicts participants' reaction time for correct recall on criterial test, which was considerably shorter for easier targets (M = 1,062 ms, SD = 774 ms) than for harder ones (M = 1,969 ms, SD = 2,069 ms), $t(48) = 2.92$, $p = .005$, $d = 0.42$. The main effect of Group was marginally significant, $F(1, 47) = 3.88$, $p = .055$, $\eta_p^2 = .08$. As shown in figure 6b, RTs were higher in the stressed group (M = 1,792, SD = 1,463) than in the control group (M = 1,176, SD = 1,624). The interaction was not significant, $F < 1$, $p = .41$.

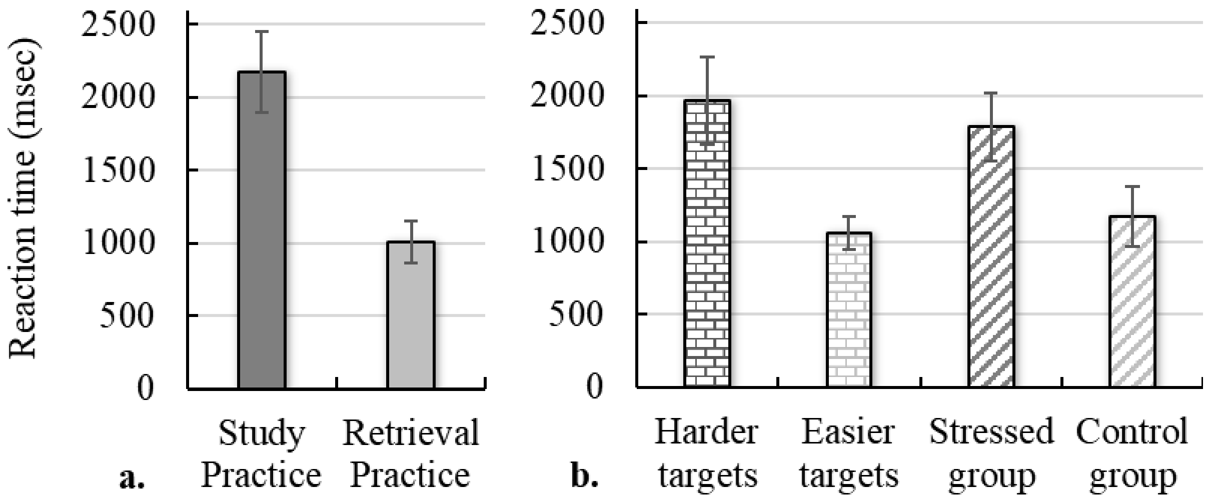

Figure 6. RT on criterial memory test for main effects of Strategy (a) and Difficulty (b).

Conditional analyses. Three different conditional analyses were conducted to better understand the influence on final performance of both effective recall during the encoding phase and effective cortisol induction before the criterial test.[2]

[2] Although conditional analyses may provide important information, we must consider that they may also raise interpretation problems regarding item selection (Karpicke et al., 2014). For example, easier items are likely to be recalled more often than harder items during practice. Consequently, easier items may benefit more from retrieval practice because they are recalled more often, not because they are intrinsically "easier".

Conditional probability of recalling on criterial test. Figure 7 depicts the probability of correct recall on the criterial test, given the number of times the same word pair was recalled during the encoding phase (for a similar procedure, see Finley, Benjamin, Hays, Bjork, & Kornell, 2011). The conditional probability for a given word pair is represented by the center of each circle, whereas the proportion of cases in each category is represented by the diameter of the circle. Data from the retrieval practice blocks were entered into a mixed logit model (based on Jaeger, 2008). Fixed effects for target difficulty and number of correct answers were entered in the model, with random-participants level intercepts. This model was then compared to empty models (i.e., which estimates if the odds of recalling and non-recalling varies between participants). The likelihood-ratio tests indicated that the addition of fixed and random terms improved the prediction for the retrieval practice model, $\chi^2(3) = 1035$, $p < .001$. Table 2 shows the model summary. For the retrieval practice model, difficulty was a significant predictor of correct final recall, with the odds of recalling an easier item on the criterial test higher than the odds of recalling a harder one (OR = 1.75). Furthermore, the odds of correct recall on the criterial test phase increased along with the number of correct answers during the encoding phase (OR = 2.35). These results corroborate the hypotheses that the benefits of the RP are partially conditioned by successful recall during the encoding phase.

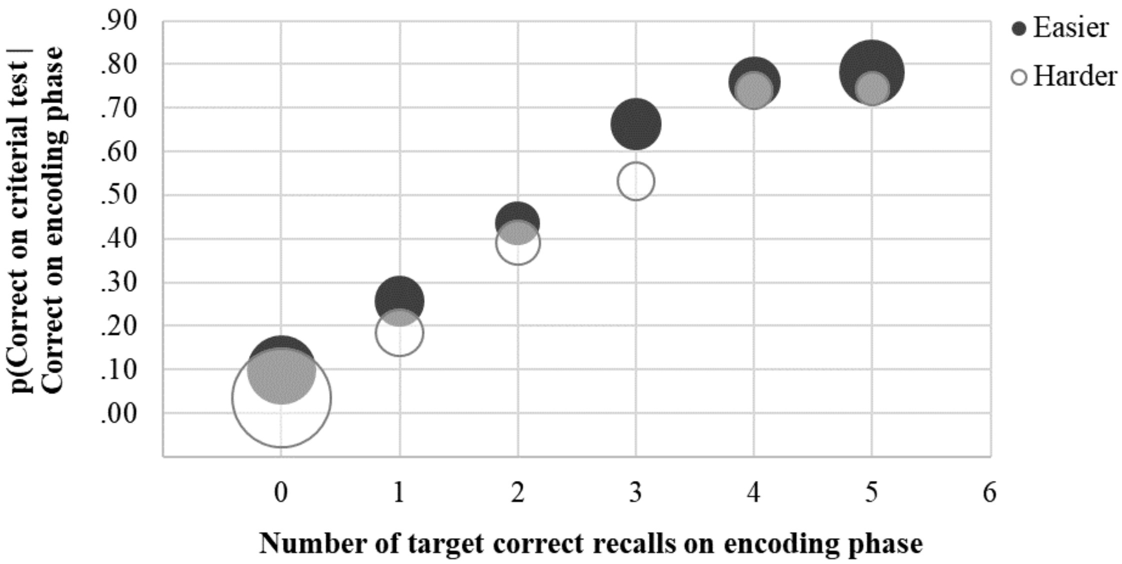

Figure 7. Conditional probabilities of successfully recalling an item on the criterial test as a function of the number of correct recalls during the encoding phase. The size of the circle indicates the number of trials included in the analysis.

Table 2. Mixed Logit Model Coefficients and Test Statistics

	β	*SE*	*Z*	*p*
Fixed effects				
Intercept	-0.84	0.11	-7.45	< .001
Number of correct recalls[a]	0.85	0.05	15.77	< .001
Difficulty	0.56	0.17	3.30	< .001
Interaction	-0.25	0.10	-2.47	0.01
Random effect	s^2			
Participants	0.32			

Note. β = log odds; *SE* = standard error; *Z* = Wald *Z* test statistic; s^2 = random effect variance. Reference categories are easier and no correct recalls. Fixed effect variables are mean centered. [a] in encoding blocks.

Cortisol responders × non-responders. Considering that the stress induction protocol successfully increased cortisol concentrations, and that the self-reported stress questionnaire (BAI3) corroborated this physiological response, we reclassified participants into two new groups based on their cortisol responses: participants who had an increase of at least 2.5 nmol/l from SC2 to SC3 were classified as "cortisol responders" ($n = 28$); the remaining participants were classified as "non-responders" ($n = 31$). This cut-off value has been recently used in a study similar to ours (Szőllősi et al., 2017) and is an adaptation from Weitzman et al. (1971). This new group division was then used for new analyses. Data from the encoding phase were entered into a 2 (Difficulty) × 5 (Block) × 2 (Group: cortisol responders or non-responders) mixed-design ANOVA, which revealed no Group effects and no Group interactions (ps > .30). The main effects of Difficulty and Block, as well as the Difficulty × Block interaction were significant (ps < .001). Importantly, there were no learning differences between groups across encoding blocks (all ps > .13). Data from the criterial test were also entered a 2 (Strategy) × 2 (Difficulty) × 2 (Group: cortisol responders or non-responders) mixed-design ANOVA, which again revealed no Group effects and no Group interactions (Fs < 1, ps > .24). The main effects of Difficulty and Block, as well as the interaction Difficulty × Block were significant (ps = .001). Because these findings were similar to the findings obtained with the original group division (control vs. stressed), no further analyses on the criterial test data were conducted using this new group division.

Conditionalized analysis. Karpicke et al. (2014) state that some researchers ignore successful retrieval on the encoding phase as a modulatory variable of the retrieval practice effect. Considering that successful retrieval on the encoding phase modulates positively the retrieval practice effect in subsequent criterial tests, and that one of the objectives of this study is to analyze this effect and its interaction with stress, we conducted an additional analysis in which the proportion of word pairs recalled on the criterial test was restricted to targets that have been recalled at least once during the encoding phase. In the original, unrestricted analysis, the proportion of targets recalled was calculated with respect to the total number of word pairs used for the retrieval practice condition (i.e., 20). In the new analysis, word pairs not recalled on practice blocks were not considered as correct responses on the criterial test results. Conditionalized data for the retrieval practice condition were entered into a 2 (Difficulty) × 2 (Group) ANOVA, which revealed a significant main effect of Difficulty, $F(1, 57) = 7.62$, $p = .008$, $\eta_p^2 = .12$, and a significant Difficulty × Group interaction, $F(1, 57) = 10.2$, $p = .002$, $\eta_p^2 = .15$, showing the same cross-over pattern obtained with the un-conditionalized data, but with larger effect sizes. The main effect of Group was not significant ($p = .72$). As shown in Figure 8, independent-samples *t*-tests revealed that (1) for easier targets, recall was lower in the stressed group ($M = .53$, $SD = .27$) than in the control group ($M = .68$, $SD = .18$), $t(51.1) = –2.43$, $p = .02$, and (2) for harder targets, recall was higher in the stressed group ($M = .56$, $SD = .34$) than in the control group ($M = .38$, $SD = .27$), $t(57) = 2.26$, $p = .03$

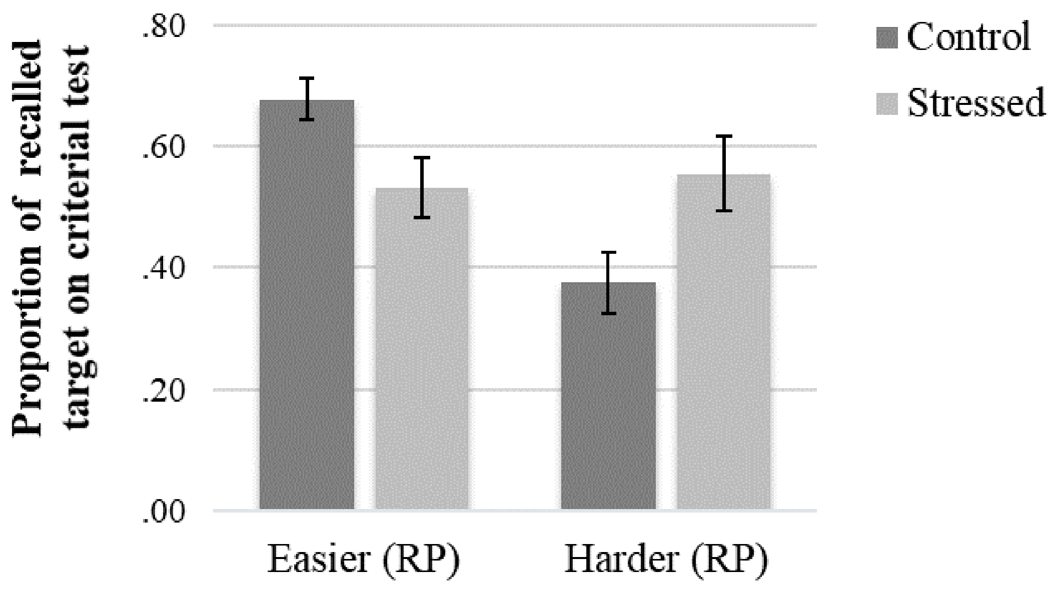

Figure 8. Effects of the stress manipulation on conditionalized data for the retrieval practice targets.

Discussion

The main objective of this study was to produce relevant evidence on (1) the RP effect and its relationship with stress, (2) item difficulty and correlated impact on memory processes, and (3) students' metacognition improvement through experience.

Our first hypothesis was that performance would be significantly higher for RP targets, which was corroborated by our results. The magnitude of the effect was higher than usual, as participants recalled 3 times more targets in the RP condition than in the SP condition (M = .36 vs. M = .12). The large effect, however, cannot be completely attributed to retrieval practice per se, as feedback was given after every recall attempt during practice blocks. Failed retrieval attempts may increase the effectiveness of subsequent encoding trials through feedback. Test-potentiated learning (Arnold & McDermott, 2012), this facilitative effect of retrieval practice on subsequent study, may thus partially account for the results.

The benefits of retrieval practice are highly modulated by successful retrieval during the encoding phase (Karpicke et al., 2014). Hence, conditionalized analyses were performed and showed that, considering only items that had been recalled at least once during learning blocks, participants recalled 4,42 times more RP targets than SP targets (M = .53 vs. M = .12). This is evidence that the success in retrieving targets during encoding modulates positively the performance in a subsequent delayed test. This result is in line with the *distribution-based bifurcation model*, according to which memory for successfully retrieved targets are highly strengthened. Feedback did not hinder the bifurcation, because even with feedback, some RP targets were not successfully retrieved on the criterial test, suggesting that they were not sufficiently boosted and stayed below the recall threshold.

Our second hypothesis was that deleterious stress effects on RP targets would not be significant. To properly test this hypothesis, two conditions had to be met: (1) the RP effect should be found, and (2) the stress manipulation (SECPT) should have to effectively induce an increase in participants' perceived stress and cortisol concentrations. Stress was successfully induced through the modified SECPT protocol, with highly significant increases in both self-reported stress (BAI3) and cortisol concentrations (*p*s = .001) in the stressed group. However, as in Szőllősi et al. (2017), no main effect of stress on memory was found. Therefore, data from this experiment does not replicate previous findings showing that stress impacts negatively the memory of participants in a between subject design (Boehringer et al., 2010; Goldfarb et al., 2017; Larrosa et al., 2017; Schwabe & Wolf, 2014).

When considering only RP targets, however, two stress-related effects emerged. First, recall of easier targets was impaired in the stressed group relative to the control group. Second, recall of harder targets was enhanced in the stressed group (on conditionalized data). These results were unexpected. One possible account is that the memory trace associated with hard targets include internal context attributes that are considered stress related by the stressed brain. In short, retrieval practice for harder items was associated with more retrieval failures (relative to easier items), and retrieval failure could be appraised as a stressful event in itself. Consequently, at retrieval and under stress, the search in memory for harder (more stressful) items may involve qualitatively different mechanisms than the search in memory for easier (less stressful) items. For harder items, stress improves recall due to the match between item-related stress and participant's current physiological state at retrieval. For easier items, stress impairs recall due to the mismatch between item-related stress and participant's state.

This account, albeit highly speculative, is partially supported by two well-known theoretical frameworks. According to the *transfer-appropriate processing hypothesis* (Morris et al., 1977), the degree of overlap between processes at encoding and retrieval may account for memory performance in several tasks (Blaxton, 1989), including retrieval-practice tasks (Veltre, Cho, & Neely, 2015). The account proposed here is analogous to transfer-appropriate processing in that the overlap between item/state at encoding and item/state at retrieval is what determines whether recall will be impaired or enhanced in the final criterial test. The proposed account is also related to the *episodic context account* (Karpicke et al., 2014) by assuming that the memory trace associated with an item also includes stress-related attributes (whether the item was successfully recalled or not during practice). Each failure to recall an item, which is more likely to occur for harder items, adds a stress-related feature to the trace composite, which is updated through feedback. During the final test and under stress, the multiple stress-related features associated to harder items help to constrain the search set, improving recall. By contrast, easier items are less likely to benefit from the item-related stress cue at test and are, therefore, more prone to the deleterious effects of stress on recall.

For SP targets, no stress-related effects on targets were found. In fact, results suggested a potential floor effect for SP targets in both groups ($M = .12$). One possible reason for this floor effect refers to the nature of the study practice task, which required typing the Portuguese translation of the Swahili word. Forcing participants to produce output by merely copying the word, without conveying meaning, is counterproductive for learning (Barcroft, 2006). Indeed, word writing may impair vocabulary learning because it diverts attentional resources from the generation of links between form and meaning (Barcroft, 2006), thus preventing deeper levels of processing at encoding (Craik & Lockhart, 1972). This is especially true if the time given to this task is short. In our experiment, participants were given 6 s to write the targets, which may not have been enough time to elaborate on meaning, consequently leading to poor performance for SP targets. The instruction to type the word already on the computer screen might have led participants to focus processing into completing the task, not into deeper processing. In Smith et al. (2016), participants in the SP group were first given 10 s to write sentences containing the to-be-learned material, generating access to meaning. Then, participants were instructed to read several times the to-be-learned information in subsequent learning blocks. This methodological difference might have increased the opportunity for meaningful associations, resulting in performance for the SP participants that was not at floor. In our experiment, this potential floor effect for SP targets may have prevented us from detecting the detrimental impact of stress on recall.

Our third hypothesis was that harder targets would be recalled at a higher rate than easier ones for RP conditions. We expected this result based on three memory theories, the *levels of processing* (Craik & Lockhart, 1972), the *retrieval effort hypothesis* (Bjork, 1975) and the *elaborative retrieval hypothesis* (Carpenter, 2009). These theories state, in different but congruent perspectives, that the harder the conditions given for recollection, the higher the memory performance, by deepening the levels of information processing, by strengthening the memory trace itself, and by forcing the generation of more semantic elaborations linked to the memory trace, respectively. By further analyzing these theories, we could notice that they refer to task difficulty, in which the procedures are manipulated (e.g., few or no cues) while attempting to retrieve previously studied information. Our data suggests that item and task difficulties play different roles in LTM. None of the theories address item intrinsic difficulty, but this was controlled for in our research, because item difficulty was an independent variable of interest. Easier targets showed significant higher recall rates than harder ones, on the criterial test for all experimental conditions: SP in both control and stressed groups (.19 vs. .04), RP in control group (.56 vs. .21), and RP in stressed group (.43 vs. .24), in line with findings from Cull and Zechmeister (1994). During practice blocks, RP easier targets were also more recalled than harder ones in both groups [control, $M_{\text{dif}} = 3.03$, $SD = 2.10$, $t(28) = 7.80$, $p < .001$, $d = 1.45$; stressed , $M_{\text{dif}} = 2.53$, $SD = 1.85$, $t(29) = 7.49$, $p < .001$, $d = 1.37$] which complicates the interpretation of the results, as it is not clear if the higher recall rate for easier over harder targets on the criterial test are partially due to this advantage on practice blocks. These results lead us to suggest that, (1) in line with the *theory of disuse* (Bjork & Bjork, 1992), more difficult to-be-learned information has lower memory retrieval strength, and (2) in line with the *bifurcation model* (Kornell et al., 2011), harder targets need more strengthening to surpass the recall threshold. We believe these two assumptions partially account for the results that easier targets were significantly more recalled than harder, as depicted in figure 5. As shown in Table 3, harder RP targets were less recalled than easier ones in the control group, as participants showed a higher recall rate for easier over harder targets when comparing the recall proportion on criterial test to the recall proportion on practice blocks (easier: .74 vs. harder: .48), the opposite results from expected according to Battig (1979), who states that more difficult items, once properly learned, will show higher recall rates in future tests. Interestingly, in the stressed group, the difference between the proportion of successful recall rates between easier and harder targets decreased dramatically (easier: .65 vs. harder .62). Furthermore, conditionalized analyses showed that stress might even revert the advantage of easier over harder items.

Some studies investigated the impact of task difficulty on memory performance (e.g., Pyc & Rawson, 2009), but the impact of item difficulty on memory performance has been less often investigated (e.g., Vaughn et al., 2013). A recent study by Minear et al. (2018), published while we were developing ours, showed that the RP effect is modulated by item difficulty and fluid intelligence. They reported similar results to ours, considering item difficulty and retrieval practice on memory performance. Item intrinsic difficulty has not been widely examined in retrieval practice experiments, especially related to stress.

Future research might enlighten this matter by addressing item difficulty as an independent variable, and also by equalizing the recall rates between easier and harder targets on the encoding phase.

Table 3. Proportion of correct RP targets recall on retrieval practice trials.

	Practice blocks		Criterial test	
Item difficulty	Easier	Harder	Easier	Harder
Un-conditionalized control	.76 (0.23)	.44 (0.26)	.56 (0.22)	.21 (0.17)
Un-conditionalized stressed	.66 (0.26)	.39 (0.29)	.43 (0.26)	.24 (0.21)
Conditionalized control[a]	1.00 (0.00)	1.00 (0.00)	.68 (0.18)	.38 (0.27)
Conditionalized stressed[a]	1.00 (0.00)	1.00 (0.00)	.53 (0.27)	.56 (0.34)

[a] Conditionalized recall rate on practice blocks were set to 1.00 because only targets recalled in this phase were considered in calculating the criterial test recall proportion.

The last hypothesis of our research was that most participants would choose SP as being more effective than RP in creating better long-term memories. This result was expected based on Karpicke et al. (2009), who showed that students chose SP over RP by answering questionnaires about which strategy they considered more effective (JOLs). Our result show that, after experiencing retrieval practice, even without any explanations about the merits of the technique, most participants said that RP would be better than SP. Participants may have noticed that retrieval practice demanded more effort and engagement and would thus be more effective than restudy. Karpicke et al. (2009) proposed that students lack awareness of the efficacy of testing as a learning strategy, but our results show the opposite, which may be due to methodological differences between studies. In our research, participants had the opportunity to use retrieval practice before they were asked about the efficacy of this technique, whereas participants were only asked about retrieval practice without having experienced the technique in Karpicke et al. (2009). In educational settings, it is important to guide students in choosing the best learning strategies, that propitiates better performance in future tests, as well as an active behavior when studying. Retrieval practice has important effects on students' metacognition, which is improved by results on performance tests and consequent more reviews (feedback) of to-be-learned information (Karpicke & Blunt, 2011).

Limitations

Some limitations are present in the current study. By exposing participants to corrective feedback, after the recall attempt during encoding phase, the retrieval practice effect is blended with the test-potentiated learning effect (Arnold & McDermott, 2012), and because of our experimental design, it is not possible to disentangle the contributions of each effect on memory performance. As stated by Pashler, Cepeda, Wixted, and Rohrer (2005), the presence of feedback after an incorrect recall attempt boosts memory performance on delayed tests. Future studies might isolate the retrieval practice effects by not providing feedback. Another limitation was the apparent floor effect for study practice targets, probably resulted from the experimental design with demanded participants to type in the translation during the practice blocks. Future research might solve this by not having participants typing in the answers. We chose to ask them to type in their answers in the study practice blocks to control for cognitive load during tasks between blocks, as retrieval practice blocks had to have the translation typed in so that measurements on the learning evolution could be analyzed.

Conclusions

Effective learning strategies are important assets that help people achieve their goals and optimize their study schedule. Retrieval practice produces better memory encoding and retrieval; hence it is a powerful means of enhancing long-term memory, which was highly corroborated by our results. The retrieval practice effect is so auspicious that Roediger, Putnam, and Smith (2011) list some RP benefits and possible applications to Education, stating that this practice can even help students when they are studying on their own.

An interesting finding in our research, that harder items were positively affected by stress when encoded by retrieval practice, but easier items were not, raises a new question about the impact of stress on memory performance related to the retrieval practice strategy. According to one view (Bjork & Bjork, 1992), both storage and retrieval strengths are boosted by retrieval practice, and performance was better for harder targets when under stress, which leads us to the assumption that stress modulates the recall threshold for these type of items. Our data shows that both easier and harder to-be-learned information benefited from retrieval practice. For educational purposes, this result is relevant in that it shows that retrieval practice may not only protect harder items from the detrimental effects of stress, but it can also boost performance of these items under stressful conditions, such as exams (e.g., SAT), presentations or interviews.

Our results also suggest that by experiencing the retrieval practice strategy, participants intuitively believed that they would obtain better results from it, compared to repeatedly studying the materials. This might play an important role in raising students' awareness of the strong benefits of retrieval practice. Schools and educators might introduce retrieval practice tasks within their methodologies to help students develop a more conscious behavior towards their decisions on how to study by themselves.

References

Abel, M., & Roediger, H. L., III. (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition, 45,* 81-92. doi:10.3758/s13421-016-0641-8

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. doi:10.3102/0034654316689306

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22,* 861-876. doi:10.1002/acp.1391

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress test: principles and practice. *Neurobiology of Stress, 6*, 113-126. doi:10.1016/j.ynstr.2016.11.001

Arnold, K., & McDermott, K. (2012). Test-potentiated learning: distinguishing between direct and indirect effects of tests. *Journal of experimental psychology. Learning, memory, and cognition, 39*. doi:10.1037/a0029199

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2015). *Memory* (2nd ed.): Psychology Press.

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research, 22*, 487-497. doi:10.1191/0267658306sr276oa

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35-56. doi:10.1111/j.1467-9922.2007.00398.x

Barcroft, J. (2015). Using retrieval opportunities to increase incidental vocabulary learning during reading: What can be done to support searners in increasing their vocabulary when reading? *Foreign Language Annals, 48*. doi:10.1111/flan.12139

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory.* (pp. 23-44). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123-144): Lawrence Erlbaum.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life.* (pp. 396-401). Oxford, England: John Wiley & Sons.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. M. a. A. Shimamura (Ed.), *Metacognition: Knowing about knowing* (First ed., pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999, 1999). *Assessing our own competence: Heuristics and illusions.* Paper presented at the Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application., Cambridge, MA, US.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory; Vol. 2: From learning processes to cognitive processes.* (pp. 35-67). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology, 64*, 417-444. doi:10.1146/annurev-psych-113011-143823

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 657-668. doi:10.1037/0278-7393.15.4.657

Boehringer, A., Schwabe, L., & Schachinger, H. (2010). A combination of high stress-induced tense and energetic arousal compensates for impairing effects of stress on memory retrieval in men. *Stress, 13*, 444-453. doi:10.3109/10253891003725256

Boyle, N. B., Lawton, C., Arkbage, K., West, S. G., Thorell, L., Hofman, D., . . . Dye, L. (2016). Stress responses to repeated exposure to a combined physical and social evaluative laboratory stressor in young healthy males. *Psychoneuroendocrinology, 63*, 119-127. doi:10.1016/j.psyneuen.2015.09.025

Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1934-1947. doi:10.1037/xlm0000427

Buchin, Z. L., & Mulligan, N. W. (2019). Divided attention and the encoding effects of retrieval. *Quaterly Journal of Experimental Psychology*, 1747021819847141. doi:10.1177/1747021819847141

Buttler, R. M., Bagci, E., Brand, H. S., den Heijer, M., Blankstein, M. A., & Heijboer, A. C. (2018). Testosterone, androstenedione, cortisol and cortisone levels in human unstimulated, stimulated and parotid saliva. *Steroids*. doi:10.1016/j.steroids.2018.05.013

Cannon, W. B. (1929). Organization for physiological homeostasis. *`Physiological reviews, 9*, 399-431. doi:10.1152/physrev.1929.9.3.399

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1563-1569. doi:10.1037/a0017021

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., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13*, 826-830. doi:10.3758/bf03194004

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633-642. doi:10.3758/BF03202713

Charmandari, E., Tsigos, C., & Chrousos, G. (2005). Endocrinology of the stress response. *Annual Review of Psychology, 67*, 259-284. doi:10.1146/annurev.physiol.67.040403.120816

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428. doi:10.1016/B978-1-4832-1446-7.50015-7

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in quantitative methods for psychology, 1*, 42-45.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition, 22*, 249-257.

Cunha, J. A. (2001). *Manual da versão em português das Escalas Beck* (Vol. 256). São Paulo.

Duchesne, A., Tessera, E., Dedovic, K., Engert, V., & Pruessner, J. C. (2012). Effects of panel sex composition on the physiological stress responses to psychosocial stress in healthy young men and women. *Biological psychology, 89*, 99-106. doi:10.1016/j.biopsycho.2011.09.009

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. doi:10.1177/1529100612453266

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paidéia, 23*, 397-406. doi:10.1590/1982-43272356201314

Everly Jr, G. S., & Lating, J. M. (2019). *A clinical guide to the treatment of the human stress response* (4th ed.). New York, NY.: Springer Nature.

Finley, J. R., Benjamin, A. S., Hays, M. J., Bjork, R. A., & Kornell, N. (2011). Benefits of accumulating versus diminishing cues in recall. *Journal of Memory and Language, 64*, 289-298. doi:10.1016/j.jml.2011.01.006

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: neurobiological mechanisms and behavioral consequences. In A. Kingstone & M. B. Miller (Eds.), *Year in Cognitive Neuroscience* (Vol. 1369, pp. 55-75). Oxford: Blackwell Science Publ.

Gagnon, S. A., Waskom, M. L., Brown, T. I., & Wagner, A. D. (2019). Stress Impairs Episodic Retrieval by Disrupting Hippocampal and Cortical Mechanisms of Remembering. *Cereb Cortex, 29*(7), 2947-2964. doi:10.1093/cercor/bhy162

Gates, A. I. (1917). *Recitation as a factor in memorizing*. New York: The Science Press.

Girdano, D., Dusek, D. E., & Everly Jr, G. S. (2012). *Controlling stress and tension* (9th ed.). Boston: Pearson Higher Ed.

Glover, J. A. (1989). The"testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*, 392.

Goldfarb, E. V., Mendelevich, Y., & Phelps, E. A. (2017). Acute stress time-dependently modulates multiple memory systems. *Journal of Cognitive Neuroscience, 29,* 1877-1894. doi:10.1162/jocn_a_01167 %M 28699809

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112. doi:10.1007/bf02289823

Groschl, M., Kohler, H., Topf, H. G., Rupprecht, T., & Rauh, M. (2008). Evaluation of saliva collection devices for the analysis of steroids, peptides and therapeutic drugs. *Journal of Pharmaceutical and Biomedical Analysis, 47*, 478-486. doi:10.1016/j.jpba.2008.01.033

Hidalgo, V., Pulopulos, M. M., Puig-Perez, S., Espin, L., Gomez-Amor, J., & Salvador, A. (2015). Acute stress affects free recall and recognition of pictures differently depending on age and sex. *Behavioral Brain Research, 292*, 393-402. doi:10.1016/j.bbr.2015.07.011

Hidalgo, V., Pulopulos, M. M., & Salvador, A. (2019). Acute psychosocial stress effects on memory performance: Relevance of age and sex. *Neurobiol Learn Mem, 157*, 48-60. doi:10.1016/j.nlm.2018.11.013

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 562-567. doi:10.1016/s0022-5371(71)80029-4

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Hupbach, A., & Fieman, R. (2012). Moderate stress enhances immediate and delayed retrieval of educationally relevant material in healthy young men. *Behavioral Neuroscience, 126*, 819-825. doi:10.1037/a0030489

Jaeger, A., Galera, C. A., Stein, L. M., & Lopes, E. J. (2016). Human memory research: Current hypotheses and new perspectives. *Estudos de Psicologia, 21*. doi:10.5935/1678-4669.20160010

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446. doi:10.1016/j.jml.2007.11.007

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112-127. doi:10.1037/0278-7393.34.1.112

Joels, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience, 10*, 459-466. doi:10.1038/nrn2632

Joels, M., Fernandez, G., & Roozendaal, B. (2011). Stress and emotional memory: a matter of timing. *Trends in Cognitive Sciences, 15,* 280-288. doi:10.1016/j.tics.2011.04.004

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*, 621-629. doi:10.1037/a0015183

Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science, 294*, 1030-1038. doi:10.1126/science.1067020

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

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

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17*, 471-479. doi:10.1080/09658210802647009

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of Learning and Motivation* (Vol. 61, pp. 237-284): Academic Press.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968. doi:10.1126/science.1152408

Keresztes, A., Kaiser, D., Kovács, G., & Racsmány, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex*, *24*, 3025-3035. doi:10.1093/cercor/bht158

Kim, J. J., & Diamond, D. M. (2002). The stressed hippocampus, synaptic plasticity and lost memories. *Nature Reviews Neuroscience, 3*, 453-462. doi:10.1038/nrn849

Kim, J. J., Song, E. Y., & Kosten, T. A. (2006). Stress effects in the hippocampus: synaptic plasticity and memory. *Stress, 9*, 1-11. doi:10.1080/10253890600678004

Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76-81. doi:10.1159/000119004

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*, 85-97. doi:10.1016/j.jml.2011.04.002

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989-998. doi:10.1037/a0015729

Larrosa, P. N. F., Ojea, A., Ojea, I., Molina, V. A., Zorrilla-Zubilete, M. A., & Delorenzi, A. (2017). Retrieval under stress decreases the long-term expression of a human declarative memory via reconsolidation. *Neurobiology of Learning and Memory, 142*(Pt A), 135-145. doi:10.1016/j.nlm.2017.03.005

Lima, M. F. R., & Buratto, L. G. (2019). *Norms for Swahili–Portuguese word pairs: Stimulus characteristics and memorability*. Manuscript submitted for publication.

Lonergan, M. H., Olivera-Figueroa, L. A., Pitman, R. K., & Brunet, A. (2013). Propranolol's effects on the consolidation and reconsolidation of long-term emotional memory in healthy participants: A meta-analysis. *Journal of Psychiatry & Neuroscience, 38*, 222-231. doi:10.1503/jpn.120111

McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371-385. doi:10.1037/0278-7393.11.2.371

McGaugh, J. L. (2000). Memory—a century of consolidation. *Science, 287*, 248-251. doi:10.1126/science.287.5451.248

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*, 1474-1486. doi:10.1037/xlm0000486

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*(5), 519-533. doi:10.1016/S0022-5371(77)80016-9

Müller, G. E., & Pilzecker, A. (1900). Experimentelle beiträge zur lehre vom gedächtnis. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane., 1*, 1-288.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2,* 325-335. doi:10.1080/09658219408258951

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *J Exp Psychol Learn Mem Cogn, 31*(1), 3-8. doi:10.1037/0278-7393.31.1.3

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods, 51*(1), 195-203.

Pulopulos, M. M., Almela, M., Hidalgo, V., Villada, C., Puig-Perez, S., & Salvador, A. (2013). Acute stress does not impair long-term memory retrieval in older people. *Neurobiology of Learning and Memory, 104,* 16-24. doi:10.1016/j.nlm.2013.04.010

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*, 437-447. doi:10.1016/j.jml.2009.01.004

Racsmany, M., Szollosi, A., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 157-166. doi:10.1037/xlm0000423

Roediger, H. L., III. (2008). Relativity of remembering: why the laws of memory vanished. *Annual Review of Psychology, 59*, 225-254. doi:10.1146/annurev.psych.57.102904.190139

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, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. doi:10.1111/j.1745-6916.2006.00012.x

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

Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In *The psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 1-36). San Diego, CA, US: Elsevier Academic Press.

Roozendaal, B., & McGaugh, J. L. (2011). Memory modulation. *Behavioral Neuroscience, 125*, 797-824. doi:10.1037/a0026187

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

Schonfeld, P., Ackermann, K., & Schwabe, L. (2014). Remembering under stress: different roles of autonomic arousal and glucocorticoids in memory retrieval. *Psychoneuroendocrinology, 39*, 249-256. doi:10.1016/j.psyneuen.2013.09.020

Schoofs, D., & Wolf, O. T. (2009). Stress and memory retrieval in women: no strong impairing effect during the luteal phase. *Behavioral Neuroscience, 123*, 547-554. doi:10.1037/a0015625

Schwabe, L., Bohringer, A., Chatterjee, M., & Schachinger, H. (2008). Effects of pre-learning stress on memory for neutral, positive and negative words: Different roles of cortisol and autonomic arousal. *Neurobiology of Learning and Memory*, *90*, 44-53. doi:10.1016/j.nlm.2008.02.002

Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology, 33*, 890-895. doi:10.1016/j.psyneuen.2008.03.001

Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neuroscience and Biobehavioral Reviews, 36*, 1740-1749. doi:10.1016/j.neubiorev.2011.07.002

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of human memory: Brain mechanisms and clinical relevance. *Biological Psychiatry, 76*, 274-280. doi:10.1016/j.biopsych.2014.03.008

Schwabe, L., & Schachinger, H. (2018). Ten years of research with the Socially Evaluated Cold Pressor Test: Data from the past and guidelines for the future. *Psychoneuroendocrinology, 92*, 155-161. doi:10.1016/j.psyneuen.2018.03.010

Schwabe, L., & Wolf, O. T. (2014). Timing matters: Temporal dynamics of stress effects on memory retrieval. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 1041-1048. doi:10.3758/s13415-014-0256-0

Selye, H. (1956). *The stress of life*. New York: Mcgraw Hill.

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin, 143*, 636-675. doi:10.1037/bul0000100

Smith, A. M., Davis, F. C., & Thomas, A. K. (2018). Criterial learning is not enough: Retrieval practice is necessary for improving post-stress memory accessibility. *Behav Neurosci, 132*(3), 161-170. doi:10.1037/bne0000240

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science*, *354*, 1046-1048. doi:10.1126/science.aah5067

Smith, A. M., Race, E., Davis, F. C., & Thomas, A. K. (2018). Retrieval practice improves item memory but not source memory in the context of stress. *Brain and Cognition*. doi:10.1016/j.bandc.2018.12.005

Smith, A. M., & Thomas, A. K. (2018). Reducing the consequences of acute stress on memory Retrieval. *Journal of Applied Research in Memory and Cognition, 7*, 219-229. doi:10.1016/j.jarmac.2017.09.007

Smith, M. A., Roediger, H. L., III, & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1712-1725. doi:10.1037/a0033569

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*, 641-656. doi:10.1037/h0063404

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*, 244-253. doi:10.3758/MC.38.2.244

Szőllősi, A., Keresztes, A., Novák, B., Szászi, B., Kéri, S., & Racsmány, M. (2017). The testing effect is preserved in stressful final testing environment. *Applied Cognitive Psychology, 31*, 615-622. doi:10.1002/acp.3363

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399. doi:10.1037/a0013082

Takahashi, H., Yoshika, M., & Yokoi, T. (2015). Validation of three automatic devices for the self-measurement of blood pressure according to the European Society of Hypertension International Protocol revision 2010: the Omron HEM-7130, HEM-7320F, and HEM-7500F. *Blood Pressure Monitoring, 20*, 92-97. doi:10.1097/mbp.0000000000000096

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175-184. doi:10.1016/s0022-5371(67)80092-6

van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage, 78*, 94-102. doi:10.1016/j.neuroimage.2013.03.071

Vaughn, K. E., Rawson, K. A., & Pyc, M. A. (2013). Repeated retrieval practice and item difficulty: Does criterion learning eliminate item difficulty effects? *Psychonomic Bulletin & Review, 20*, 1239-1245. doi:10.3758/s13423-013-0434-z

Veltre, M. T., Cho, K. W., & Neely, J. H. (2015). Transfer-appropriate processing in the testing effect. *Memory, 23*, 1229-1237.

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: implications for the classroom. *npj Science of Learning, 1*. doi:10.1038/npjscilearn.2016.11

Weitzman, E. D., Fukuchima, D., Nogeire, C., Roffwarg, H., Gallagher, T. F., & Hellman, L. (1971). Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *The Journal of Clinical Endocrinology & Metabolism, 33*, 14-22. doi:10.1210/jcem-33-1-14

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*, 240-245. doi:10.1111/j.1467-9280.1992.tb00036.x

Wolf, O. T. (2017). Stress effects on learning and memory in humans. In *Stress: Neuroendocrinology and Neurobiology* (Vol. 2, pp. 309). San Diego, CA, US: Elsevier Academic Press.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*, 515-525. doi:10.1016/j.tins.2011.06.006

Zoladz, P. R., Peters, D. M., Kalchik, A. E., Hoffman, M. M., Aufdenkampe, R. L., Woelke, S. A., . . . Talbot, J. N. (2014). Brief, pre-learning stress reduces false memory production and enhances true memory selectively in females. *Physiology & Behavior, 128*, 270-276. doi:10.1016/j.physbeh.2014.02.028

Appendix A: Experimental Stimuli Used

Table A1

Parameters of Swahili–Portuguese Word Pairs Used in Retrieval Practice Blocks

Difficulty	WP number	Swahili	Portuguese	English[a]	Memorability[b]
harder	01	inda	malícia	spite	0.29
	02	iktisadi	economia	economy	0.26
	03	yamini	juramento	oath	0.25
	04	hamira	levedura	yeast	0.18
	05	ruba	sanguessuga	leech	0.28
	06	hariri	seda	silk	0.28
	07	lozi	amêndoa	almond	0.22
	08	ziwa	lago	lake	0.28
	09	bustani	jardim	garden	0.26
	10	nanga	ancora	anchor	0.27
easier	11	elimu	ciência	science	0.47
	12	nabii	profeta	prophet	0.49
	13	punda	burro	donkey	0.7
	14	usingizi	sono	sleep	0.47
	15	wingu	nuvem	cloud	0.57
	16	lulu	perola	pearl	0.73
	17	pipa	barril	barrel	0.75
	18	goti	joelho	knee	0.51
	19	mbwa	cachorro	dog	0.65
	20	pombe	cerveja	beer	0.65

[a] Original English word normed for Nelson and Dunlosky (1994).

[b] Memorability was computed as average proportion of participants that correctly recalled items over three blocks of tests (see, Lima & Buratto, 2019).

Appendix A: Experimental Stimuli Used (continued)

Table A2

Parameters of Swahili–Portuguese Word Pairs Used in Study Practice Blocks

Difficulty	WP number	Swahili	Portuguese	English[a]	Memorability[b]
harder	21	ladha	sabor	flavor	0.14
	22	hadithi	história	story	0.28
	23	utenzi	poema	poem	0.23
	24	jeraha	ferida	wound	0.27
	25	handaki	trincheira	trench	0.14
	26	ankra	fatura	invoice	0.22
	27	samadi	estrume	manure	0.25
	28	ambo	cola	glue	0.25
	29	bahasha	envelope	envelope	0.21
	30	lango	portão	gate	0.28
easier	31	roho	alma	soul	0.67
	32	chama	sociedade	society	0.65
	33	wasaa	lazer	leisure	0.59
	34	dafina	tesouro	treasure	0.63
	35	malkia	rainha	queen	0.73
	36	buu	larva	maggot	0.53
	37	chura	sapo	frog	0.47
	38	nafaka	milho	corn	0.51
	39	godoro	colchão	mattress	0.47
	40	nyanya	tomate	tomato	0.75

[a] Original English word normed for Nelson and Dunlosky (1994).

[b] Memorability was computed as average proportion of participants that correctly recalled items over three blocks of tests (see, Lima & Buratto, 2019).

Appendix B: Additional Analysis

According to the automatization account of the RP effect (Racsmany et al., 2018), the benefits derived from retrieval practice stem from the automatization it promotes, which leads to lower involvement of other cognitive functions (e.g., attention) and consequent faster processing. The authors compared the benefits obtained from retrieval practice to those of skill learning in several situations, showing that empirical evidence from different studies corroborate this assumption. Some of the benefits observed in both skill learning and retrieval practice include better performance in tasks involving divided attention (Buchin & Mulligan, 2017, 2019) and proactive interference (Szpunar, McDermott, & Roediger, 2008). The key dependent variable accounted to support the theory is the RT of retrieval, that is, retrieval practiced items are recalled faster than restudied ones. Racsmany et al. (2018) also showed that data from two functional magnetic resonance imaging (fMRI) studies (Keresztes, Kaiser, Kovács, & Racsmány, 2014; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013) supporting their theory of automatization because, during retrieval, participants had (1) increased activity in the striatum and (2) decreased fronto-parietal activity for retrieval practiced items, both being typical characteristics observed in skill learning studies. Although it was not our goal when designing this experiment, our results are congruent with the automatization account theory proposed by Racsmany et al. (2018). Our results show that RTs for recall attempts during practice blocks decreased following a power function, as shown in Figure 9.

Appendix B: Additional Analysis (continued)

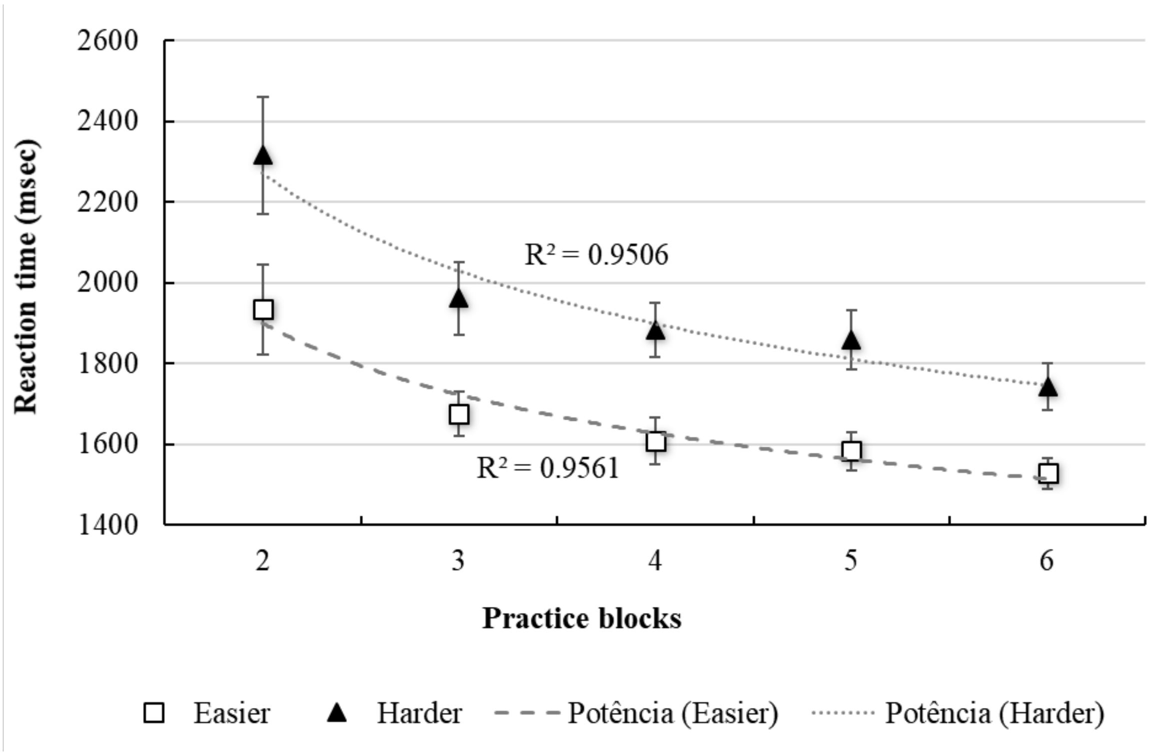

Figure 9. Fitted power function to averaged RT on practice blocks.

Appendix C: Ethics Committee Approval

UNB - INSTITUTO DE CIÊNCIAS HUMANAS E SOCIAIS DA UNIVERSIDADE

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Como a estratégia de estudo chamada de Prática de Recuperação pode codificar memórias de longo prazo que sejam resistentes aos efeitos nocivos do estresse, em especial do hormônio cortisol, no momento de se lembrar ou evocar o conteúdo estudado anteriormente.

Pesquisador: CARLOS EDUARDO DANTAS DA COSTA LAGE

Área Temática:

Versão: 2

CAAE: 80059617.2.0000.5540

Instituição Proponente: Instituto de Psicologia -UNB

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.558.544

Apresentação do Projeto:

PROJETO DE MESTRADO - IP - UNB

A EFICÁCIA DA PRÁTICA DE RECUPERAÇÃO NA CRIAÇÃO DE MEMÓRIAS RESISTENTES AO ESTRESSE

O objetivo deste projeto, portanto, é avaliar se prática de recuperação pode gerar memórias de longo-prazo mais resistentes aos efeitos nocivos do cortisol para a memória. Sessenta participantes adultos, do sexo masculino, serão expostos a 40 pares de palavras (Suaíli-Português). Metade dos pares serão aprendidos via prática de recuperação, e a outra metade via releitura. Uma semana depois, os participantes farão um teste final de memória. Antes do teste, metade dos participantes passará por um protocolo de indução ao estresse (SECPT) e a outra metade, não. A hipótese é que participantes no grupo submetido a estresse lembrarão menos pares de palavras que participantes do grupo sem estresse, mas que esse prejuízo será menor para palavras que foram originalmente aprendidas via prática de recuperação.

Hipótese: Quando a estratégia de estudos chamada Prática de Recuperação for utilizada para a codificação da memória, a evocação das informações estudadas são protegidas de eventuais prejuízos decorrentes da presença do hormônio cortisol no organismo, que é uma consequência

Endereço: CAMPUS UNIVERSITÁRIO DARCY RIBEIRO - FACULDADE DE DIREITO - SALA BT 03/1 (Ao lado da Direção)
Bairro: ASA NORTE **CEP:** 70.910-900
UF: DF **Município:** BRASILIA
Telefone: (61)3107-1592 **E-mail:** cep_chs@unb.br

Appendix C: Ethics Committee Approval (continued)

UNB - INSTITUTO DE CIÊNCIAS HUMANAS E SOCIAIS DA UNIVERSIDADE

Continuação do Parecer: 2.558.544

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

BRASILIA, 22 de Março de 2018

Assinado por:
Érica Quinaglia Silva
(Coordenador)

Endereço: CAMPUS UNIVERSITÁRIO DARCY RIBEIRO - FACULDADE DE DIREITO - SALA BT 03/1 (Ao lado da Direção)
Bairro: ASA NORTE **CEP:** 70.910-900
UF: DF **Município:** BRASILIA
Telefone: (61)3107-1592 **E-mail:** cep_chs@unb.br

Appendix D: Consent Form for the Control Group

Universidade de Brasília

Termo de Consentimento Livre e Esclarecido - TCLE

Convidamos você a participar do projeto "A eficácia da prática de recuperação na criação de memórias resistentes ao estresse", sob a responsabilidade de Carlos Eduardo Dantas da Costa Lage, aluno de mestrado no Instituto de Psicologia (UnB). O objetivo desta pesquisa é verificar se diferentes técnicas de memorização podem reduzir os efeitos prejudiciais do estresse na memória.

Sua participação se dará por meio de dois encontros de 45 minutos, o primeiro para estudar uma lista de palavras, e o segundo, 7 dias após o primeiro, para a realização de um teste de memória. Os dois encontros serão realizados no LIPSI (Laboratório Integrado de Pós-Graduação e Pesquisa Experimental em Psicologia com Humanos), no Instituto de Psicologia (UnB, campus Darcy Ribeiro), na parte da tarde.

No dia do teste, você realizará um procedimento que envolve colocar uma de suas mãos dentro de um recipiente com água a temperatura ambiente durante 2 minutos. Esse procedimento será realizado antes do teste de memória. Esse protocolo é amplamente utilizado em pesquisas em diversos países, pois é uma maneira rápida e econômica de gerar respostas fisiológicas de estresse. Nessa fase do experimento, também serão realizadas algumas medições de pressão arterial e coletas de amostras de saliva para posterior análise em laboratório. Caso queira receber os resultados das análises laboratoriais, um email deve ser fornecido ao pesquisador. Os dados de contato não serão utilizados para nenhum outro fim e não serão divulgados de nenhuma forma.

Ao participar desse estudo, você estará contribuindo para o desenvolvimento de técnicas de estudos mais eficazes. Você pode se recusar a responder qualquer questão que lhe traga constrangimento, pode se recusar a ser filmado e pode se recusar a realizar o procedimento de indução de estresse. Além disso, você pode desistir de participar da pesquisa a qualquer momento sem que isso lhe traga qualquer prejuízo. Sua participação é voluntária e não haverá pagamento por sua colaboração.

Os resultados da pesquisa serão divulgados em seminários e publicações científicas. Seus dados serão mantidos sob sigilo, sendo removidas quaisquer informações que permitam identificá-lo. Os dados obtidos nessa pesquisa ficarão sob a guarda do pesquisador por cinco anos. Após esse período, eles serão destruídos.

Caso tenha qualquer dúvida em relação à pesquisa, por favor telefone para Carlos Eduardo no telefone 061-99944-2999 ou envie um e-mail para cadu@aluno.unb.br.

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da Universidade de Brasília. Dúvidas com relação à assinatura deste TCLE ou sobre seus direitos como participante da pesquisa podem ser esclarecidas pelo telefone (61) 3107-1592 ou pelo email cep_chs@unb.br. O CEP/CHS se localiza na

Página **1** de **2**

Appendix D: Consent Form for the Control Group (continued)

Universidade de Brasília

Faculdade de Direito, na sala BT 03/1, Campus Universitário Darcy Ribeiro, Universidade de Brasília, Asa Norte.

Caso concorde em participar, pedimos que assine este documento que foi elaborado em duas vias. Uma via ficará com o pesquisador responsável, e a outra, com você.

Dados do participante:

Nome completo: ____

Data de Nascimento: ____/____/________ CPF: ___________________________

Telefone "Whatsapp": (______) _______________________________

Email: ____

____

Assinatura do participante

____

Assinatura do pesquisador responsável

Brasília, ___ de __________de____.

Participante NÚMERO

Página **2** de **2**

Appendix E: Consent Form for the Experimental Group

Universidade de Brasília

Termo de Consentimento Livre e Esclarecido - TCLE

Convidamos você a participar do projeto "A eficácia da prática de recuperação na criação de memórias resistentes ao estresse", sob a responsabilidade de Carlos Eduardo Dantas da Costa Lage, aluno de mestrado no Instituto de Psicologia (UnB). O objetivo desta pesquisa é verificar se diferentes técnicas de memorização podem reduzir os efeitos prejudiciais do estresse na memória.

Sua participação se dará por meio de dois encontros de 45 minutos, o primeiro para estudar uma lista de palavras, e o segundo, 7 dias após o primeiro, para a realização de um teste de memória. Os dois encontros serão realizados no LIPSI (Laboratório Integrado de Pós-Graduação e Pesquisa Experimental em Psicologia com Humanos), no Instituto de Psicologia (UnB, campus Darcy Ribeiro), na parte da tarde.

No dia do teste, você realizará um procedimento de indução de estresse fisiológico. Nesse procedimento, você irá colocar uma mão dentro de um recipiente com água gelada durante 2 minutos para indução de resposta de estresse fisiológico no organismo. Esse procedimento será realizado antes do teste de memória. A água gelada deverá gerar desconforto, mas o efeito é transitório. Esse protocolo é amplamente utilizado em pesquisas em diversos países, pois é uma maneira rápida e econômica de gerar respostas fisiológicas de estresse. Você será filmado durante a realização desse procedimento. Nessa fase do experimento, também serão realizadas algumas medições de pressão arterial e coletas de amostras de saliva para posterior análise em laboratório. Caso queira receber os resultados das análises laboratoriais, um email deve ser fornecido ao pesquisador. Os dados de contato não serão utilizados para nenhum outro fim e não serão divulgados de nenhuma forma.

Ao participar desse estudo, você estará contribuindo para o desenvolvimento de técnicas de estudos mais eficazes. Você pode se recusar a responder qualquer questão que lhe traga constrangimento, pode se recusar a ser filmado e pode se recusar a realizar o procedimento de indução de estresse. Além disso, você pode desistir de participar da pesquisa a qualquer momento sem que isso lhe traga qualquer prejuízo. Sua participação é voluntária e não haverá pagamento por sua colaboração.

Os resultados da pesquisa serão divulgados em seminários e publicações científicas. Seus dados serão mantidos sob sigilo, sendo removidas quaisquer informações que permitam identificá-lo. Os dados obtidos nessa pesquisa ficarão sob a guarda do pesquisador por cinco anos. Após esse período, eles serão destruídos.

Caso tenha qualquer dúvida em relação à pesquisa, por favor telefone para Carlos Eduardo no telefone 061-99944-2999 ou envie um e-mail para cadu@aluno.unb.br.

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da Universidade de Brasília. Dúvidas com relação à assinatura deste

Página **1** de **2**

Appendix E: Consent Form for the Experimental Group (continued)

Universidade de Brasília

TCLE ou sobre seus direitos como participante da pesquisa podem ser esclarecidas pelo telefone (61) 3107-1592 ou pelo email cep_chs@unb.br. O CEP/CHS se localiza na Faculdade de Direito, na sala BT 03/1, Campus Universitário Darcy Ribeiro, Universidade de Brasília, Asa Norte.

Caso concorde em participar, pedimos que assine este documento que foi elaborado em duas vias. Uma via ficará com o pesquisador responsável, e a outra, com você.

Dados do participante:

Nome completo: ____

Data de Nascimento: ____/____/________ CPF: ____________________________

Telefone "Whatsapp": (______) ________________________________

Email: ____

____

Assinatura do participante

____

Assinatura do pesquisador responsável

Brasília, ___ de __________de____.

Participante NÚMERO

Página **2** de **2**

Appendix F: Checklist Form Used During Sessions

Departamento de Processos Psicológicos Básicos
Instituto de Psicologia, Universidade de Brasília

A eficácia da Prática de Recuperação na criação de memórias resistentes ao estresse.

CHECKLIST PARA SESSÃO EXPERIMENTAL 01

Data: ____/____/________

Hora início: _____:_____

Hora término: _____:_____

Participante NÚMERO

Nome do participante:____

CPF: ______________________________ WhatsApp: (____) ______________________

Assinatura do participante: ______________________________________

☐ Leitura e assinatura do TCLE (termo de consentimento livre e esclarecido);

☐ Marque o programa em PsychoPy utilizado: [A] [B];

☐ COLETA de Saliva ____;

☐ Pressão: Sistólica ______ mm/Hg, Diastólica ______ mm/Hg Pulso ______ bpm.

☐ Observações: ____

____

____

____

Appendix F: Checklist Form Used During Sessions (continued)

Departamento de Processos Psicológicos Básicos
Instituto de Psicologia, Universidade de Brasília

CHECKLIST PARA SESSÃO EXPERIMENTAL 02

Data: ____/____/2018 Hora início: _____:_____ Hora término: _____:_____

Assinatura do participante: ____

☐ PsychoPy programa: session_2_questionnaires;

☐ COLETA de Saliva ____;

☐ Pressão: Sistólica ______ mm/Hg, Diastólica ______ mm/Hg Pulso ______ bpm;

☐ Marque grupo experimental SECPT: [ ]estressado [ ]controle;

☐ Cold Pressor: Temperatura ______°C Tempo submerso: _____ min ______ seg;

☐ Vídeo: the_office_episode (grupo estressado: matemática e Cold Pressor ligado);

☐ PsychoPy programa: session_2_final_memory_test;

☐ COLETA de Saliva ____;

☐ Pressão: Sistólica ______ mm/Hg, Diastólica ______ mm/Hg Pulso ______ bpm;

☐ Leitura e assinatura do formulário de Debriefing e do recibo.

☐ Observações: ____

____

Appendix G: Debriefing Form

Departamento de Processos Psicológicos Básicos
Instituto de Psicologia, Universidade de Brasília

A eficácia da Prática de Recuperação na criação de memórias resistentes ao estresse.

DEBRIEFING

Obrigado pela sua participação. A hipótese principal do estudo é: em momentos de estresse, pessoas que estudam utilizando a técnica conhecida como Prática de Recuperação têm melhores resultados em se lembrar de itens guardados na memória do que pessoas que estudaram por outras técnicas, como a releitura do material. Para que o estresse fosse realmente induzido no participante, utilizamos elementos como dor física (mão imersa na água gelada), avaliação social (câmera gravando o participante para um comitê avaliar suas reações faciais à dor), e imprevisibilidade (solicitar uma contagem de 100 para trás, de 7 em 7, inesperadamente no meio do vídeo "The office". A câmera não estava realmente gravando, ou seja, foi utilizada somente para gerar a resposta fisiológica do estresse no participante e nenhuma imagem ou áudio foram salvos. O grupo controle do experimento somente colocou a mão em água na temperatura ambiente, sem os fatores de avaliação social e imprevisibilidade.

RECIBO

Eu, ____, CPF ________________

declaro haver recebido, como reembolso de despesas alimentares dos dois dias das sessões experimentais, o total de R$20,00 (vinte reais).

Participante NÚMERO

Brasília, _______ de ______________ de 2018.

____

Assinatura do participante