*Applied Cognitive Psychology, Appl. Cognit. Psychol.* (2017)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.3363

# The Testing Effect is Preserved in Stressful Final Testing Environment

ÁGNES SZŐLLŐSI[1]*, ATTILA KERESZTES[2], BÁLINT NOVÁK[3], BARNABÁS SZÁSZI[4], SZABOLCS KÉRI[1,5] and MIHÁLY RACSMÁNY[1,6]

[1]*Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary*
[2]*Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany*
[3]*Department of General Psychology, Pázmány Péter Catholic University, Budapest, Hungary*
[4]*Institute of Psychology, Eötvös Loránd University, Budapest, Hungary*
[5]*National Institute of Psychiatry and Addictions, Nyírő Gyula Hospital, Budapest, Hungary*
[6]*Research Group on Frontostriatal Disorders, Hungarian Academy of Sciences, Budapest, Hungary*

*Summary: Previous studies have shown that retrieval practice leads to better long-term memory than additional study of a material (a phenomenon termed the testing effect). In this study, we compared the effectiveness of these learning strategies when the final test occurs under stress (such as in an exam). Participants studied word pairs; then half of the material was repeatedly studied, whereas the other half was repeatedly tested. Following a 7-day delay, participants were exposed to either a psychosocially stressful situation or a control task, followed by an associative recall task that tested memory for all items. Multiple measures were used to assess stress levels: emotional state assessments as well as assays of salivary cortisol and alpha-amylase levels. Results are in favour of the ecological validity of retrieval-based learning. Participants recalled more retested items than restudied items regardless of being exposed to a stressful situation and the hormonal (cortisol) response to stress.*

To find ways to make our memories more resistant to forgetting is a fundamental goal not only in memory research but also in our everyday life. Testing, or retrieval practice, has been a learning technique of special interest because previous studies have pointed out the active role of retrieval in memory retention (Roediger & Karpicke, 2006a; Tulving, 1967). The testing effect refers to the long-term retention benefit that results from retrieving a material when compared with restudying it after initial learning (Roediger & Karpicke, 2006a, 2006b).

The efficiency of retrieval-based learning in laboratory settings is underscored by a large number of studies (for a review, see, e.g. Roediger & Butler, 2011); furthermore, there is strong empirical evidence for the beneficial effect of testing in educational practice (e.g. McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014; Roediger, Agarwal, McDaniel, & McDermott, 2011). Results of these studies are in favour of the ecological validity of retrieval-based learning as they show that the testing effect pertains with different materials and test formats widely used in educational settings (Roediger & Butler, 2011; Roediger & Karpicke, 2006b). However, a study by Hinze and Rapp (2014) showed that when restudy–retest practice occurred under high pressure in an educational setting, it eliminated the testing effect (Hinze & Rapp, 2014). Importantly, the typical testing situations in everyday life are usually exams, where despite the considerable stress experienced, it is especially important to recall the previously acquired knowledge as accurately as possible. To investigate the relationship between stress and the effectiveness of different learning strategies seems to be especially important, because it is known that stress and stress hormones clearly influence episodic memory retrieval (i.e. the conscious recollection of events).

Experiencing stressful situations triggers the activation of the hypothalamic–pituitary–adrenal axis and the sympathetic nervous system (Mason, 1968; O'Connor, O'Halloran, & Shanahan, 2000). Free salivary cortisol levels and salivary alpha-amylase (sAA) activity are reliable markers of the activations of the hypothalamic–pituitary–adrenal axis (Kirschbaum & Hellhammer, 1994) and the sympathetic nervous system, respectively (Nater et al., 2006). In a typical experiment on the relationship between stress, stress hormones and human memory, participants are presented with a memory paradigm following cortisone administration or stress exposure. Most studies found that elevated cortisol levels before retrieval impair the retrieval of various learning materials (e.g. Kirschbaum, Wolf, May, Wippich, & Hellhammer, 1996; de Quervain, Rozendaal, Nitsch, McGaugh, & Hock, 2000; Wolf et al., 2001; Schwabe & Wolf, 2014; for reviews, see Lupien, Maheu, Tu, Fiocco, & Schramek, 2007; Wolf, 2009) and autobiographical events (Buss, Wolf, Witt, & Hellhammer, 2004; Schlosser et al., 2010). A wide range of glucocorticoid receptors can be found in the hippocampus, which structure is known to play a key role in episodic memory (McEwen, 2008). Results of functional neuroimaging studies suggest that cortisol decreases the activations of the hippocampus resulting in impaired ability to access (episodic) information previously obtained (de Quervain et al., 2003; Oei et al., 2007; Pruessner et al., 2008).

It is important to note that in all the previously mentioned studies on stress-related memory, participants had no chance to practise the learning material in a systematic way after initial learning. Therefore, it is unclear how previous findings can be generalized to differences in memory for materials learnt with different strategies. Because there is a great intersubject variability in the way individuals react to stressors (Miller, Plessow, Kirschbaum, & Stalder, 2013), it seems to

*Correspondence to: Ágnes Szőllősi, Department of Cognitive Science, Budapest University of Technology and Economics, Egry József Street 1, 1111 Budapest, Hungary.
E-mail: aszollosi@cogsci.bme.hu

be also important to investigate cortisol effects on memory and their relationship with different learning strategies.

In this study, we aimed to investigate the long-term effectiveness of restudy and retest practice when the final memory test occurred under stress. Following the initial learning of paired associates, participants practised the word pairs either by rereading the material (restudy condition) or by cued recall (retest condition). Feedback was given for each item during practice, because the inclusion of feedback is suggested to use in educational practice as it improves the beneficial effect of practice (e.g. Butler, Karpicke, & Roediger, 2007; Butler & Roediger, 2008). Following a 7-day retention interval, subjects were exposed to either psychosocial stressors or a non-stressful control task. Finally, participants' memory was tested for all word pairs they studied previously.

Importantly, to assess stress levels, multiple measures were used: assays of salivary cortisol and alpha-amylase levels as well as emotional state assessments. Stress levels were assessed three times: immediately before the initial learning phase as well as immediately before and after the stress inducing (or the control) task.

It has been recently demonstrated that repeated retrieval practice decreases the involvement of attentional control (Mulligan & Picklesimer, 2016) and of attentional control-related brain regions (Keresztes, Kaiser, Kovács, & Racsmány, 2014; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013) and increases the level of automatization of recall (Racsmány, Szőllősi, & Bencze, 2017). An important attribute of automatization is that memories become more resistant to various disturbing effects (Logan, 1988). Therefore, we could assume that the automatization of retrieval through repeated retrieval practice is an important protective factor against the negative effects of acute stress.

## MATERIALS AND METHODS

### Participants

Sixty-seven Hungarian undergraduate students (native Hungarian speakers) participated in the experiment. One participant's baseline salivary cortisol level (45 nmol/l) was more than three standard deviations away from the mean of the sample ($M$ = 17.6 nmol/l, $SD$ = 9.1); therefore, this participant was considered as an outlier. Further, three participants were excluded from the sample, because they did not provide enough saliva for cortisol analysis. We analysed the results of 63 participants (28 men and 35 women; age range: 19–27 years; $M_{age}$ = 21.3 years, $SD$ = 1.7). Participants were randomly assigned into one of the two experimental groups. On experimental day 2, they were exposed to either a stressful situation (stress group; $n$ = 30; 13 men; $M_{age}$ = 21.4 years, $SD$ = 1.5) or a control task (control group; $n$ = 33; 15 men; $M_{age}$ = 21.2 years, $SD$ = 1.8).

Subjects were recruited at different universities in Budapest, Hungary, and received either extra course credits or money for their participation (type of compensation was equally distributed between the experimental groups). The study was approved by the Ethical Committee of the Budapest University of Technology and Economics, Hungary.

### Memory task

Participants were presented with a computer-controlled learning paradigm, while seated at approximately 70 cm from a computer display. The experiment was performed using PRESENTATION® software (version 14.3, www.neurobs.com). We used word pairs as stimuli, because in tasks using paired associates as stimuli, it is easy to provide feedback to the participants and to control for the time interval between study and test (Roediger & Karpicke, 2006b). Stimuli were 40 neutral Swahili–Hungarian word pairs translated from Nelson and Dunlosky (1994).

The memory task consisted of two main phases, separated by a 7-day retention interval: a learning phase (initial learning and practice) and a final test phase. In the initial learning phase, participants were presented with all word pairs in random order [5000 milliseconds per word pair; inter-stimulus interval (ISI): 500 milliseconds] on the computer screen, with the Swahili word on the left and its Hungarian equivalent on the right, and were instructed to memorize as many word pairs as they could. Because the retention benefit of testing increases as a function of the number of practice trials (e.g. Rawson & Dunlosky, 2011; Vaughn & Rawson, 2011; Wiklund-Hörnqvist, Jonsson, & Nyberg, 2014), our subjects practised the word pairs in six cycles immediately after the initial learning phase. Word pairs were randomly assigned into a restudy (20 word pairs) or a retest condition (20 word pairs). Each practice cycle consisted of a restudy, a retest and a feedback block (the order of the restudy and retest blocks varied randomly across the learning cycles). In the *restudy* blocks, participants saw 20 Swahili words together with their Hungarian meanings in random order (8000 milliseconds per word pair; ISI: 500 milliseconds). In the *retest* blocks, 20 Swahili words were presented in random order on the computer screen, and participants were instructed to type the Hungarian meanings of the words using the keyboard of the computer. They had a maximum of 8000 milliseconds to complete one word pair. Based on the results of earlier experiments using retrieval practice manipulations, delayed feedback seems to be more beneficial to memory retention than immediate feedback (e.g. Butler et al., 2007; Butler & Roediger, 2008) probably because memory retention is better when the stimulus presentation is spaced or distributed (Dempster, 1989). Therefore, in our experiment (following some previous studies, e.g. Keresztes et al., 2014), feedback was not given after each trial, but at the end of each practice cycle. In the feedback blocks, all 40 word pairs were presented randomly for the participants (1500 milliseconds per word pair; ISI: 500 milliseconds). In order to eliminate the effect of self-testing during the 7-day retention interval, participants were informed that the purpose of the second experimental session would be to examine social cognition.

Seven days after the first experimental session, participants' memory for all 40 word pairs was tested in the final test phase. Swahili words were presented in random order,

and participants were asked to type their Hungarian equivalents. They had a maximum of 8000 milliseconds to complete one word pair.

### Stress and control procedure

Because testing is typically a socially stressful situation in educational settings, we decided not to use a physically but a socially stressful procedure. Therefore, participants in the stress group were exposed to the Trier Social Stress Test (TSST; Kirschbaum, Pirke, & Hellhammer, 1993), which was developed to induce moderate psychosocial stress in humans in laboratory settings. Participants had 5 minutes to prepare for giving a 5-minute speech in front of two observers (a male and a female) who had been introduced as a board of experts in non-verbal behaviour. Following the 5-minute preparatory phase and the 5-minute speech, participants were given a 5-minute arithmetic task. We told them that audio and video recordings would also be made for later analysing their behaviour (no recording was actually made). Participants in the control group were exposed to a standardized control version of the TSST (Het, Rohleder, Schoofs, Kirschbaum, & Wolf, 2009), which was developed to be as similar as possible to the original stress-inducing procedure but without any socially stressful components (observers as well as audio and video recorders).

### Saliva sampling and cortisol/alpha-amylase analyses

Saliva samples were collected from each participant three times: once on the first experimental day and twice on the second experimental day. Samples were collected using Eppendorf Safe-Lock Tubes (1.5 ml) and were kept at room temperature until the end of the experimental sessions, and then at −10 °C until analysis (for a maximum of 12 weeks). Free salivary cortisol concentrations and sAA activity were determined by Salimetrics immunoassay.

### Subjective assessment

Following each saliva sampling, participants completed the Hungarian version of the state form of the State-Trait Anxiety Inventory (STAI-s; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983; Hungarian: Sipos, Sipos, & Spielberger, 1994) including 20 questions on the current level of anxiety. Participants were asked to rate the items of the questionnaire on a 4-point scale (1 = *not at all* and 4 = *very much so*). Immediately after the completion of the STAI-s, participants were asked to fill out the Hungarian version of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988; Hungarian: Gyollai, Simor, Köteles, & Demetrovics, 2011) including two sets of items on positive (10 items) and negative affects (PANAS-n; 11 items). Participants were instructed to complete the PANAS on their current affective state. They rated the items on a 5-point scale ranging from 1 = *not at all* to 5 = *extremely*. Finally, subjects rated how stressful they were just then (0 = *not at all* and 100 = *very stressful*).

### General procedure

Participants were asked to refrain from meal, alcohol, caffeine, smoking and physical exercise 2 hours prior to the experimental sessions in order to eliminate any effect of these factors on salivary cortisol levels and on sAA activity. Experimental sessions were run between 16:00 and 20:00 in order to avoid any interference with the cortisol circadian cycle (for a review, see, e.g. Clow, Hucklebridge, Stalder, Evans, & Thorn, 2010) as well as the circadian rhythm-dependent change in daily sAA activity (Rohleder, Nater, Wolf, Ehlert, & Kirschbaum, 2004).

For the experimental procedure, see Figure 1. The first experimental session was preceded by a 5-minute preparatory phase in order to familiarize the participants to the situation and to minimize the effect of stress-inducing factors (e.g. new environment) in this initial phase of the experiment. During this preparatory phase, participants gave written informed consent and completed a preliminary questionnaire including questions on demographic data and any known neurological and psychiatric disorders. The preparation was followed by a stress assessment (saliva sampling and subjective rating), and then subjects participated in the learning phase of the memory task (initial learning and the six practice cycles).

The second session (7 days after the first session) was also preceded by a preparatory phase that was followed by a stress assessment (saliva sampling and subjective rating). Immediately after the stress assessment, participants were exposed to either the TSST or the control procedure. Although most earlier studies used a delay between the stress/control manipulation and testing (in order to reach the cortisol peak), following Hupbach and Fieman (2012), we aimed to minimize the time interval between the stress-inducing procedure and the final test phase of the memory task in order to preserve the context of the stressful situation as much as possible (following more an exam-like situation). Therefore, the TSST/Control-TSST was immediately followed by the third stress assessment (saliva sampling and subjective rating) that took 3 to 5 minutes and then, again with no delay, the final test of the memory task.

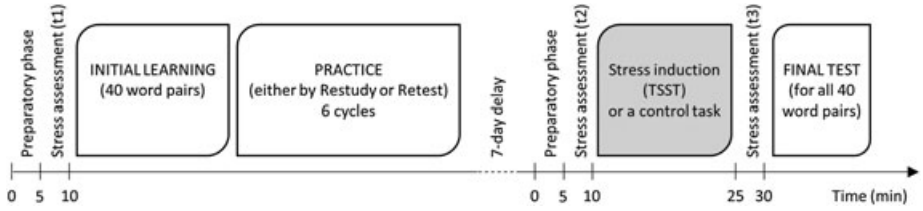


Figure 1. Experimental procedure. Stress assessment: saliva sampling and subjective rating (state form of the State-Trait Anxiety Inventory, the positive and negative affect schedule and a subjective stress scale); $t_1$, immediately before initial learning; $t_2$, immediately before stress/control manipulation; $t_3$, immediately after stress/control manipulation; TSST, Trier Social Stress Test

## RESULTS

### Stress versus control procedure

*The success of the stress induction*

Different types of measurements were used to assess the success of the stress induction (determination of cortisol and sAA levels from saliva and emotional state assessments; Figure 2). For all measurements, we conducted a mixed-design analysis of variance (ANOVA) with stress exposure (stress/control) as a between-subjects variable and time ($t_1/t_2/t_3$) as a within-subjects variable (where $t_1$ = immediately before initial learning, $t_2$ = immediately before the stress/control task and $t_3$ = immediately after the stress/control task).

For sAA activity, the ANOVA established the success of the stress induction. We found a significant Time × Stress exposure interaction [$F(2, 58) = 44.36$, $p < .001$, $\eta_p^2 = .43$]. The contrast analysis showed that sAA activity in the stress group increased from $t_2$ to $t_3$ [$F(1, 27) = 40.67$, $p < .001$, $\eta_p^2 = .60$]. For cortisol data, a similar pattern of results emerged [interaction between time and stress exposure: $F(2, 61) = 3.00$, $p = .056$, $\eta_p^2 = .05$], with stressed subjects' cortisol levels showing a trend-level increase from $t_2$ to $t_3$ [$F(1, 29) = 3.60$, $p = .068$, $\eta_p^2 = .11$]. The lower effect sizes for cortisol levels than for sAA activity could be because sAA activity reaches its peak considerably faster than salivary cortisol levels (Nater et al., 2005). Note that we did not use a delay between stress and saliva sampling in order to preserve the context of the stressful situation.

Mean ratings on the PANAS-n, STAI-s and the subjective stress scale showed similar patterns. We found significant Time × Stress exposure interactions [PANAS-n: $F(1, 61) = 6.53$, $p = .002$, $\eta_p^2 = .10$; STAI-s: $F(1, 61) = 9.15$, $p < .001$, $\eta_p^2 = .13$; subjective stress: $F(1, 61) = 7.23$, $p = .001$, $\eta_p^2 = .11$]. Independent samples *t*-tests established that scores differed between the stress and the control groups at $t_3$ [PANAS-n: $t(61) = 3.04$, $p = .004$, $d = 0.78$; STAI-s: $t(61) = 3.22$, $p = .002$, $d = 0.82$; subjective stress: $t(61) = 2.81$, $p = .007$, $d = 0.72$]. Similarly to results obtained for cortisol and amylase levels, the contrast analysis showed that stressed subjects received higher scores at $t_3$ than at $t_2$ [PANAS-n: $F(1, 29) = 0.34$, $p = .028$, $\eta_p^2 = .16$; STAI-s: $F(1, 29) = 16.21$, $p < .001$, $\eta_p^2 = .36$; subjective stress: $F(1, 29) = 20.04$, $p < .001$, $\eta_p^2 = .41$]. Altogether, this pattern of results confirmed that the stress induction was successful.

*Memory performance*

In order to test whether initial test performance differed between the groups at the end of the first (learning) phase of the experiment, we compared recall success in the last retest cycle between the stress and the control condition (stress group: $M = 85.7\%$, $SD = 17.1$; control group: $M = 91.4\%$, $SD = 12.5$). We found no significant difference between the groups [$t(61) = 0.94$, $p = .352$, $d = 0.24$].

A mixed-design ANOVA was conducted on recall rate in the final test phase of the memory task with study condition (restudy/retest) as a within-subjects variable and stress exposure (stress/control) as a between-subjects variable. For recall rates, a significant main effect of study condition was found [$F(1, 61) = 44.01$, $p < .001$, $\eta_p^2 = .42$]. Post-hoc analyses established that participants recalled more retested items than restudied items in both groups [stress group: $t(29) = 5.66$, $p < .001$, $d = 1.03$; control group: $t(32) = 3.80$, $p < .001$, $d = 0.66$; Figure 3(A)]. Stress exposure had no effect on recall performance—neither the main effect of stress exposure [$F(1, 61) = 0.88$, $p = .351$, $\eta_p^2 = .01$] nor the interaction between the independent variables (Stress exposure × Study condition) was significant [$F(1, 61) = 1.41$, $p = .240$, $\eta_p^2 = .02$].

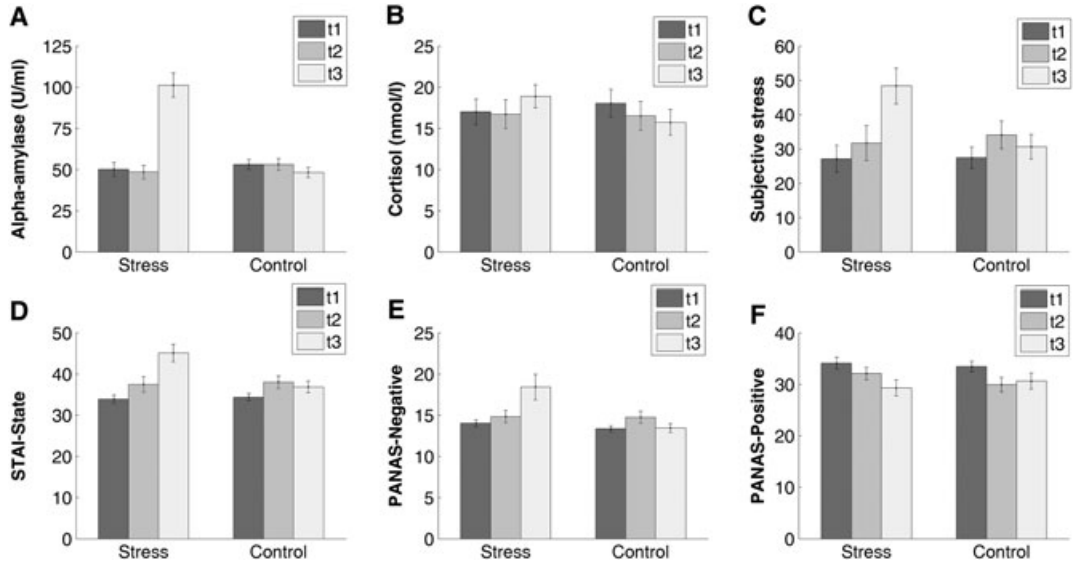


Figure 2. The impact of stress exposure as shown by different measures. An increase could be seen from $t_2$ to $t_3$ in stressed subjects' (A) sAA activity ($p < .001$), (B) salivary cortisol levels ($p = .068$), (C) subjective stress levels ($p < .001$), (D) STAI-state scores ($p < .001$) and (E) PANAS-negative scores ($p = .028$). (F) No significant group difference was found in participants' positive affective state as measured by the PANAS-positive subscale. Error bars represent the standard error of the mean. PANAS, Positive and Negative Affect Schedule; STAI, State-Trait Anxiety Inventory; $t_1$, immediately before initial learning; $t_2$, immediately before stress/control manipulation; $t_3$, immediately after stress/control manipulation

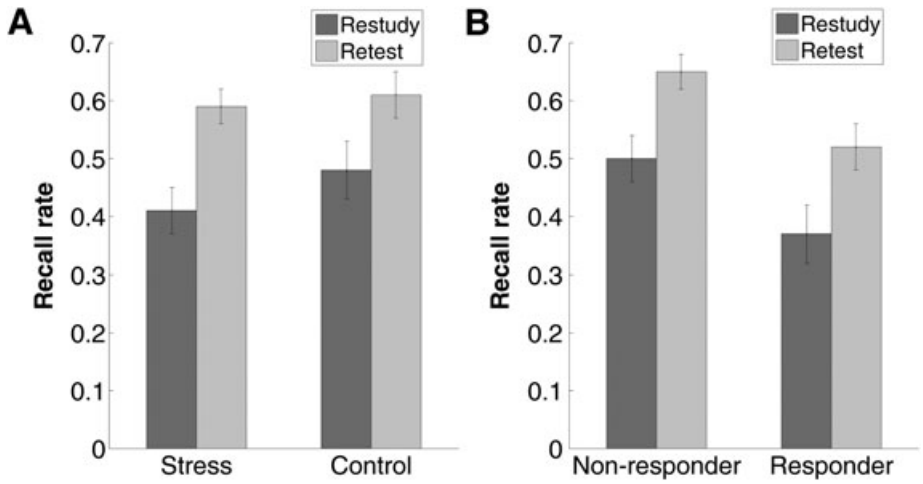


Figure 3. Recall rates in the final test phase of the memory task. Comparison of recall rates for the restudied and the retested word pairs (A) between the stress group and the control group and (B) between cortisol responders and non-responders. Error bars represent the standard error of the mean

In brief, despite the successful stress induction, stress (immediately before retrieval) had no effect on memory performance. Furthermore, retrieval practice led to better long-term memory retention than rereading both in the stress and the control groups.

## Cortisol responders versus non-responders

### *Composition of the cortisol responder and the non-responder group*

On a post-hoc basis, we classified each participant into either a cortisol responder or a non-responder group (irrespective of whether they were exposed to the stress-inducing task or the control procedure). The cut-off point we used was adapted from Weitzman et al. (1971). Participants in the responder group ($n$ = 24; 10 men; also member of the stress group: $n$ = 16) showed an increase of at least 2.5 nmol/l in their cortisol levels immediately after the stress/control procedure when compared with their own baseline levels immediately before the stress exposure or the control task. The remaining 39 participants (18 men; also member of the stress group: $n$ = 14) were assigned into the non-responder group.

### *Memory performance*

The 2 × 2 ANOVA (Study condition × Cortisol response) revealed a significant main effect of study condition [$F(1, 61) = 40.00$, $p < .001$, $\eta_p^2 = .40$] and of cortisol response [$F(1, 61) = 6.00$, $p = .017$, $\eta_p^2 = .09$] indicating that repeated retrieval led to superior long-term memory performance when compared with rereading and that subjects who showed a cortisol response performed worse (in general) on the long-term recall test than non-responders. The Study condition × Cortisol response interaction was not significant [$F(1, 61) = 1.00$, $p > .99$, $\eta_p^2 < .01$].

Post-hoc analyses showed that cortisol responders recalled fewer items both in the restudy [$t(61) = 2.05$, $p = .045$, $d = 0.52$] and the retest conditions [$t(61) = 2.44$, $p = .018$, $d = 0.62$]. Despite this, recall rate for the retested word pairs was higher than for the restudied items in the responder group [$t(23) = 3.65$, $p = .001$, $d = 0.74$] and also in the non-responder group [$t(38) = 5.50$, $p < .001$, $d = 0.88$; Figure 3(B)].

Because non-responders showed better memory than cortisol responders in the final test phase, we tested on a post-hoc basis whether the two groups differed in their initial test performance (i.e. during the retest cycles of the practice phase; Figure 4). We conducted a mixed-design ANOVA with a within-subjects factor of cycle (1–6) and a between-subjects factor of cortisol response (responders/non-responders). Whereas the Cycle × Cortisol response interaction was not significant [$F(1, 61) = 0.63$, $p = .679$, $\eta_p^2 = .01$], cycle and cortisol response had main effects on recall rate [$F(1, 61) = 419.38$, $p < .001$, $\eta_p^2 = .87$ and $F(1, 61) = 4.73$, $p = .034$, $\eta_p^2 = .07$, respectively]. This pattern of results indicate that non-responders showed better initial learning performance than cortisol responders (irrespective the presence of any stressors in this phase of the experiment).

In a following analysis, we compared the proportion of correct initial test items on final test between cortisol responders ($M$ = 61.0%, $SD$ = 18.0) and non-responders ($M$ = 70.1%, $SD$ = 17.2), and we found a significant difference between the groups [$t(61) = 2.00$, $p = .050$, $d = 0.51$]. These findings indicate that besides their relatively low initial learning performance, cortisol responders showed higher forgetting rate when compared with the non-responder group.

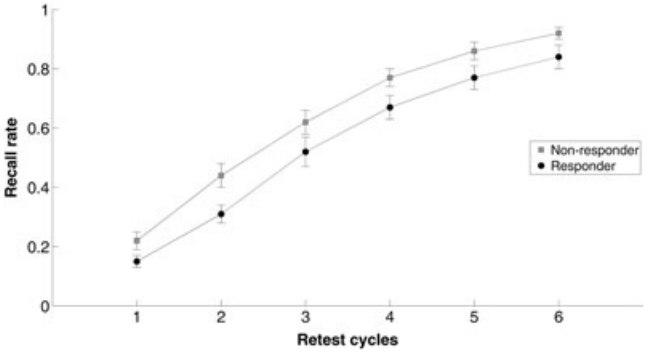


Figure 4. Recall rates of cortisol responders and non-responders in the six retest cycles of the practice phase. Error bars represent the standard error of the mean

In sum, cortisol responders recalled fewer word pairs than non-responders in the final test phase—because of their difference in initial learning success and forgetting rate. Nevertheless, retrieval practice seemed to be a more efficient strategy than rereading even if individuals showed an increase in their cortisol levels (responders) and even if they did not (non-responders).

## DISCUSSION

The main objective of our study was to test whether the long-term benefit of retrieval practice (retest) compared with rereading (restudy) is insensitive to stress exposure. Participants showed an increase in their negative affective state (as measured by the PANAS-n subscale, the state form of the STAI and a subjective stress scale) and sAA activity immediately after stress exposure when compared with the control group—confirming the success of the stress induction. Memory for the retested word pairs was better than for the restudied word pairs irrespective of whether participants were presented with psychosocial stressors or not.

Importantly, although stress treatment had no significant effect on our participants' memory performance, the composition of a cortisol responder and a non-responder group provided an opportunity to test cortisol effects on memory. Cortisol responders performed worse (in general) on the final memory test than non-responders because of their difference in initial learning performance and forgetting rate. Individuals with cortisol response showed lower learning rates than non-responders during practice, and cortisol response was associated with higher long-term forgetting. These findings suggest that cortisol response did not have an exclusive effect on retrieval. It seems that there is a general relationship between individual differences in cortisol responsiveness and the success of learning irrespective of the presence of any stressors. In fact, previous studies have shown that there is a relationship between glucocorticoid levels, the volume of the hippocampus and memory performance (e.g. Lindauer, Olff, Van Meijel, Carlier, & Gersons, 2006; Lupien et al., 1998; Travis et al., 2016; for reviews, see, e. g. Frodl & O'Keane, 2013; Kim & Diamond, 2002). Accordingly, in our study, greater cortisol responsiveness was associated with reduced learning rate might be due to differential hippocampal functioning in cortisol responders versus non-responders. Nevertheless, despite their relatively low initial learning performance and high forgetting rate, retrieval-based learning was more beneficial for long-term retention than restudy practice in cortisol responders as well.

In a recent study, Smith, Floerke, and Thomas (2016) investigated the impact of acute psychosocial stress on the testing effect. Similarly to our results, the authors found a strong testing effect and no stress effects on memory immediately after the stress exposure. However, in this study, stress had a detrimental impact on the recall of the restudied items but not on the recall of the retested items 20 minutes after stress induction. In our study, cortisol response had a negative effect not only on the recall of restudied items but also on the recall of retested items. A possible solution of this contradiction is an important methodological difference between the present study and the study of Smith et al. (2016): in our study, feedback was administered during practice. It has been widely demonstrated that feedback during practice enhances the retention benefit of testing (e.g. Butler et al., 2007, Butler & Roediger, 2008; Kang, McDermott, & Roediger, 2007). In our experiment, recall improved steadily during retrieval practice as a consequence of feedback, indicating that some new learning occurred during practice. Thus, elevated cortisol levels during final recall acted on memories acquired through the combination of study (as a consequence of feedback) and retest. However, and importantly, cortisol did not influence the beneficial effect of retrieval practice over restudy as suggested by the fact that we found no difference in the magnitude of the testing effect between responders and non-responders. A possible aim of future studies should be to compare the effect of cortisol response (and stress) on retrieval-practised memories using paradigms with and without feedback during testing.

In another study, Hinze and Rapp (2014) have demonstrated that when retrieval practice occurred under high pressure, the testing effect has become eliminated. At a first glance, the findings of Hinze and Rapp (2014) seem to contradict our results and also the results of Smith et al. (2016). However, these contradictory findings may stem from important methodological differences between these studies. First, Hinze and Rapp (2014) did not manipulate and measure the level of acute stress. Instead, they manipulated the pressure on task performance and the importance of the test during practice. In contrast, our study and also the study of Smith et al. (2016) manipulated the level of acute stress only before the final test and held constant the importance of the task. These factors (such as the motivation level of the participants and the level of anxiety) could have differential effects. Considering the results of Hinze and Rapp (2014), it could be the case that the testing effect is unaffected by acute stress only when the source of the stressor is not related to the importance of the task and to the motivational level of the participants.

Another important methodological difference is the amount of retrieval practice. Hinze and Rapp (2014) applied a single quiz-like test on complex scientific texts. In contrast, in our study, participants repeatedly practised paired-associate items in a cued recall task, whereas the participants of Smith et al. (2016) repeatedly retrieved a list of items in a free recall situation. It has been demonstrated that repeated retrieval practice increases the level of automatization of recall (Racsmány et al., 2017). Because memories become more resistant to various disturbing effects as a result of automatization (Logan, 1988), we could assume that the automatization of retrieval following retrieval practice is an important protective factor against the negative effects of acute stress.

From an applied perspective, our finding seems to be especially significant, because there are several psychosocially stressful situations in everyday life (e.g. school exams and job interviews), when it is crucial whether the previously acquired knowledge is accessible or not. The possible aim of future research should be to investigate whether our results on the relationship between cortisol levels and the testing effect can be generalized to real-world educational settings. Furthermore, in fact, one session of testing more closely

represents a typical schooling situation. Because there are examples of studies demonstrating the retention benefit of testing over restudy following only one practice block (see, e.g. Hinze & Rapp, 2014), another aim of future studies should be to investigate the relationship between stress and the testing effect using one practice session. Finally, it should be mentioned as a caveat that our conclusion that acute stress does not harm the effect of retrieval practice is based on a failure to reject the null hypothesis. Nevertheless, according to our results, it seems that the long-term retention benefit of retrieval practice in comparison with rereading (i.e. the testing effect) is insensitive to any aspects of the stress protocol we used and also to the hormonal response to stress.

## ACKNOWLEDGEMENTS

We thank Petra Kovács, Dóra Molnár-Bakos and Bertalan Polner for their help in data collection. This work was supported by the KTIA NAP Grant (ID: 13-2-2014-0020). Ágnes Szőllősi is supported by the ÚNKP-16-3-III New National Excellence Program of the Ministry of Human Capacities, Hungary.

## REFERENCES

Buss, C., Wolf, O. T., Witt, J., & Hellhammer, D. H. (2004). Autobiographical memory impairment following acute cortisol administration. *Psychoneuroendocrinology*, *29*, 1093–1096. https://doi.org/10.1016/j.psyneuen.2003.09.006.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*, 273–281. https://doi.org/10.1037/1076-898X.13.4.273.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition*, *36*, 604–616. https://doi.org/10.3758/MC.36.3.604.

Clow, A., Hucklebridge, F., Stalder, T., Evans, P., & Thorn, L. (2010). The cortisol awakening response: More than a measure of HPA axis function. *Neuroscience and Biobehavioral Reviews*, *35*, 97–103. https://doi.org/10.1016/j.neubiorev.2009.12.011.

de Quervain, D. J., Henke, K., Aerni, A., Treyer, V., McGaugh, J. L., Berthold, T., … Hock, C. (2003). Glucocorticoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe. *European Journal of Neuroscience*, *17*, 1296–1302. https://doi.org/10.1046/j.1460-9568.2003.02542.x.

de Quervain, D. J., Rozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience*, *3*, 313–314. https://doi.org/10.1038/73873.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309–330. https://doi.org/10.1007/BF01320097.

Frodl, T., & O'Keane, V. (2013). How does the brain deal with cumulative stress? A review with focus on developmental stress, HPA axis function and hippocampal structure in humans. *Neurobiology of Disease*, *52*, 24–37. https://doi.org/10.1016/j.nbd.2012.03.012.

Gyollai, A., Simor, P., Köteles, F., & Demetrovics, Z. (2011). Psychometric properties of the Hungarian version of the original and the short form of the Positive and Negative Affect Schedule (PANAS). *Neuropsychopharmacologia Hungarica*, *13*(2), 73–79.

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. T. (2009). Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. *Psychoneuroendocrinology*, *34*, 1075–1086. https://doi.org/10.1016/j.psyneuen.2009.02.008.

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, *28*, 597–606. https://doi.org/10.1002/acp.3032.

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. https://doi.org/10.1037/a0030489.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). The format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558. https://doi.org/10.1080/09541440601056620.

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. https://doi.org/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. https://doi.org/10.1038/nrn849.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*, *19*, 313–333. https://doi.org/10.1016/0306-4530(94)90013-2.

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

Kirschbaum, C., Wolf, O. T., May, M., Wippich, W., & Hellhammer, D. H. (1996). Stress- and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life Sciences*, *58*, 1475–1483. https://doi.org/10.1016/0024-3205(96)00118-X.

Lindauer, R. J. L., Olff, M., Van Meijel, E. P. M., Carlier, I. V. E., & Gersons, B. P. R. (2006). Cortisol, learning, memory, and attention in relation to smaller hippocampal volume in police officers with posttraumatic stress disorder. *Biological Psychiatry*, *59*, 171–177. https://doi.org/10.1016/j.biopsych.2005.06.033.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527. https://doi.org/10.1037/0033-295X.95.4.492.

Lupien, S. J., de Leon, M., de Santi, S., Convit, A., Tarshish, C., Nair, N. P. V., … Meaney, M. J. (1998). Cortisol levels during human aging predict hippocampal atrophy and memory deficits. *Nature Neuroscience*, *1*, 69–73. https://doi.org/10.1038/271.

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition*, *65*, 209–237. https://doi.org/10.1016/j.bandc.2007.02.007.

Mason, J. W. (1968). A review of psychoneuroendocrine research on the sympathetic–adrenal medullary system. *Psychosomatic Medicine*, *30*, 631–653.

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

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*, 3–21. https://doi.org/10.1037/xap0000004.

McEwen, B. S. (2008). Central effects of stress hormones in healthy and disease: Understanding the protective and damaging effects of stress and stress mediators. *European Journal of Pharmacology*, *583*, 174–185. https://doi.org/10.1016/j.ejphar.2007.11.071.

Miller, R., Plessow, F., Kirschbaum, C., & Stalder, T. (2013). Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: Evaluation of salivary cortisol pulse detection in panel designs. *Psychosomatic Medicine*, *75*, 832–840. https://doi.org/10.1097/PSY.0000000000000002.

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

Nater, U. M., La Marca, R., Florin, L., Moses, A., Langhans, W., Koller, M. M., & Ehlert, U. (2006). Stress-induced changes in human salivary alpha-amylase activity—Associations with adrenergic activity. *Psychoneuroendocrinology*, *31*, 49–58. https://doi.org/10.1016/j.psyneuen.2005.05.010.

Nater, U. M., Rohleder, N., Gaab, J., Berger, S., Jud, A., Kirschbaum, C., & Ehlert, U. (2005). Human salivary alpha-amylase reactivity in a psychosocial stress paradigm. *International Journal of Psychophysiology*, *55*, 333–342. https://doi.org/10.1016/j.ijpsycho.2004.09.009.

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

O'Connor, T. M., O'Halloran, D. J., & Shanahan, F. (2000). The stress response and the hypothalamic–pituitary–adrenal axis: From molecule to melancholia. *Quarterly Journal of Medicine*, *93*, 323–333. https://doi.org/10.1093/qjmed/93.6.323.

Oei, N. Y. L., Elzinga, B. M., Wolf, O. T., de Ruiter, M. B., Damoiseaux, J. S., Joost, P., … Rombouts, S. A. R. B. (2007). Glucocorticoids decrease hippocampal and prefrontal activation during declarative memory retrieval in young men. *Brain Imaging and Behavior*, *1*, 31–41. https://doi.org/10.1007/s11682-007-9003-2.

Pruessner, J. C., Dedovic, K., Khalili-Mahani, N., Engert, V., Pruessner, M., Buss, C., … Lupien, S. (2008). Deactivation of the limbic system during acute psychosocial stress: Evidence from positron emission tomography and functional magnetic resonance imaging studies. *Biological Psychiatry*, *63*, 234–240. https://doi.org/10.1016/j.biopsych.2007.04.041.

Racsmány, M., Szőllősi, Á., & Bencze, D. (2017). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (Epub ahead of print, 29 Jul 2017). DOI: https://doi.org/10.1037/xlm0000423.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*, 283–302. https://doi.org/10.1037/a0023956.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395. https://doi.org/10.1037/a0026252.

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

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

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

Rohleder, N., Nater, U. M., Wolf, J. M., Ehlert, U., & Kirschbaum, C. (2004). Psychosocial stress-induced activation of alpha-amylase: An indicator of sympathetic activity. *Annals of the New York Academy of Sciences*, *1032*, 258–263. https://doi.org/10.1196/annals.1314.033.

Schlosser, N., Wolf, O. T., Fernando, S. C., Riedesel, K., Otte, C., Muhtz, C., … Wingenfeld, K. (2010). Effects of acute cortisol administration on autobiographical memory in patients with major depression and healthy controls. *Psychoneuroendocrinology*, *35*, 316–320. https://doi.org/10.1016/j.psyneuen.2009.06.015.

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

Sipos, K., Sipos, M., & Spielberger, C. D. (1994). A State and Trait Anxiety Inventory (STAI) magyar változata (The Hungarian version of the State and Trait Anxiety Inventory [STAI]). In F. Mérei, & F. Szakács (Eds.), *Pszichodiagnosztikai Vademecum*, (pp. 123 – 148). Budapest: Nemzeti Tankönyvkiadó.

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

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Travis, S. G., Coupland, N. J., Hegadoren, K., Silverstone, P. H., Huang, Y., Carter, R., … Malykhin, N. V. (2016). Effects of cortisol on hippocampal subfields volumes and memory performance in healthy control subjects and patients with major depressive disorder. *Journal of Affective Disorders*, *201*, 34–41. https://doi.org/10.1016/j.jad.2016.04.049.

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. https://doi.org/10.1016/S0022-5371(67)80092-6.

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

Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science*, *22*, 1127–1131. https://doi.org/10.1177/0956797611417724.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063.

Weitzman, E. D., Fukushima, 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 and Metabolism*, *33*, 14–22. https://doi.org/10.1210/jcem-33-1-14.

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*, 1–7. https://doi.org/10.1111/sjop.12093.

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research*, *1293*, 142–154. https://doi.org/10.1016/j.brainres.2009.04.013.

Wolf, O. T., Convit, A., McHugh, P. F., Kandil, E., Thorn, E. L., De Santi, S., … de Leon, M. J. (2001). Cortisol differentially affects memory in young and elderly men. *Behavioral Neuroscience*, *115*, 1002–1011. https://doi.org/10.1037/0735-7044.115.5.1002.