Journal of Experimental Psychology: Learning, Memory, and Cognition

Retrieval Practice Makes Procedure From Remembering: An Automatization Account of the Testing Effect

Mihály Racsmány, Ágnes Szőllősi, and Dorottya Bencze

Online First Publication, June 29, 2017. http://dx.doi.org/10.1037/xlm0000423

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000423

RESEARCH REPORT

Retrieval Practice Makes Procedure From Remembering: An Automatization Account of the Testing Effect

Mihály Racsmány
Hungarian Academy of Sciences and Budapest University of Technology and Economics

Ágnes Szőllősi and Dorottya Bencze
Budapest University of Technology and Economics

The "testing effect" refers to the striking phenomenon that repeated retrieval practice is one of the most effective learning strategies, and certainly more advantageous for long-term learning, than additional restudying of the same information. How retrieval can boost the retention of memories is still without unanimous explanation. In 3 experiments, focusing on the reaction time (RT) of retrieval, we showed that RT of retrieval during retrieval practice followed a power function speed up that typically characterizes automaticity and skill learning. More important, it was found that the measure of goodness of fit to this power function was associated with long-term recall success. Here we suggest that the automatization of retrieval is an explanatory component of the testing effect. As a consequence, retrieval-based learning has the properties characteristic of skill learning: diminishing involvement of attentional processes, faster processing, resistance to interference effects, and lower forgetting rate.

Keywords: automatization, memory, retrieval practice, skill learning, testing effect

One of the most solid cornerstones of any theory of human memory has long been the distinction between the encoding and retrieval phases of memory processing, the former being responsible for storing, whereas the latter is responsible for access of learned information (e.g., Atkinson, 1996). This concept was heavily challenged by results showing that repeated retrieval practice itself is an effective learning strategy even more successful in promoting long-term learning than repeated study, a benefit widely referred to as the testing effect (Roediger & Butler, 2011; Roediger & Karpicke, 2006).

In a typical experimental procedure of the testing effect literature, following an initial learning phase, participants took part either in a retrieval practice or in a repeated study task with the initially studied items (Karpicke, Lehman, & Aue, 2014). The final retrieval of all items could be either a few minutes or days after the practice phase. The most robust finding is that items practiced through retrieval show decreased forgetting and have a long-term benefit relative to study practice (Roediger & Butler, 2011; Roediger & Karpicke, 2006). Most of the studies found an interaction between the lengths of delay between practice and final recall (minutes or days) and the form of practice (repeated retrieval or study), showing a short-term advantage of repeated study and a long-term retrieval practice benefit (Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003; but see Karpicke et al., 2014). Prominent and partly conflicting accounts of the testing effect proposed that high strengths of memory traces are because of the effort required to retrieve a specific item, or assumed that every act of retrieval adds new semantically appropriate cues or new temporal/contextual features to retrieved memories (Carpenter, 2009; Karpicke et al., 2014; Kornell, Bjork, & Garcia, 2011). However, the long-term benefit of retrieval-based learning is still without unanimous explanation.

It is interesting that having a closer look on the key findings of the literature, one can find strong similarities between retrieval-based learning and skill learning. The information acquired through repeated retrieval is characterized with different attributes in comparison with repeatedly studied information. For instance, information learned by repeated retrieval is (a) more resistant to interference effects (Racsmány & Keresztes, 2015; Szpunar, McDermott, & Roediger, 2008), (b) shows a lower forgetting rate following weeks or months (Roediger & Butler, 2011; Roediger & Karpicke, 2006), and (c) remains accessible in multitasking situations where attentional processes are heavily loaded (Mulligan & Picklesimer, 2016). Compellingly, the previous properties of

Mihály Racsmány, Frontostriatal System Research Group, Hungarian Academy of Sciences, and Department of Cognitive Science, Budapest University of Technology and Economics; Ágnes Szőllősi and Dorottya Bencze, Department of Cognitive Science, Budapest University of Technology and Economics.

This work was supported by the KTIA NAP 13 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. We thank Ágnes Lukács for her useful comments on the manuscript. We thank Dóra Molnár-Bakos and Miklós Marián for their help in data collection.

Correspondence concerning this article should be addressed to Mihály Racsmány, Department of Cognitive Science, Budapest University of Technology and Economics, Egry Jozsef u. 1, 1111 Budapest, Hungary. E-mail: racsmany@cogsci.bme.hu

retrieval-based learning are also characteristics of skill learning (Kuhl, Dudukovic, Kahn, & Wagner, 2007; Newell & Rosenbloom, 1981; Schneider & Chein, 2003; Squire & Zola, 1996).

Moreover, it is also known that spaced initial retrieval practice produces greater memory benefits than does massed initial retrieval (Jacoby, 1978; Whitten & Bjork, 1977). Similarly, distributed practice in simple and complex skill learning is also superior to massed learning (Lee & Genovese, 1988). Another interesting finding is that retrieval practice with relatively infrequent and weak retrieval cues produces better memory than the same practice with strong and frequent cues (Carpenter, 2009). Similarly, it was found that reduced frequency of knowledge of results enhances motor skill learning (Winstein & Schmidt, 1990). The number of practice trials is important in both retrieval-based learning and skill learning, because it was found that one or two practice trials are less beneficial than a higher amount of repeated practice (Hanawalt, 1937; Logan, 1988).

Along with these similarities, two recent functional magnetic resonance imaging (fMRI) studies drew attention to another dependent variable that should be relevant to all account of human memory (Ratcliff & McKoon, 2000), the RT of retrieval (Keresztes, Kaiser, Kovács, & Racsmány, 2014; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). Both studies found that retested items were retrieved faster than were restudied ones, both 20 min (Keresztes et al., 2014) and 7 days (Keresztes et al., 2014; van den Broek et al., 2013) following practice. These results were in line with previous studies that found decreased retrieval latencies during selected retrieval practice (Keresztes & Racsmány, 2013; Román, Soriano, Gómez-Ariza, & Bajo, 2009), retrieval of semantic facts (Pirolli & Anderson, 1985), and list recall (Lehman, Smith, & Karpicke, 2014). Van den Broek and colleagues (2013) scanned participants during retrieval/restudy practice and found increased activity in the striatal cortex, in the thalamus, and in the associative cortex during retrieval practice, a finding resembling patterns typically observed in skill learning studies (Raichle et al., 1994). Keresztes and colleagues (2014) scanned participants during a cued-recall task either 20 min or 1 week following retest/restudy practice and found decreased control network activity for retested items in comparison with restudied items, with no change in activation level after a 7-day delay. These results suggest that retrieval practice produces faster retrieval at the final test, and that this acceleration is associated with increased basal ganglia and decreased control network activities, again, a typical finding in skill learning literature (Hikosaka, Nakamura, Sakai, & Nakahara, 2002; Kuhl et al., 2007; Newell & Rosenbloom, 1981; Schneider & Chein, 2003).

Altogether, these attributes of test-enhanced learning could point to the hypothesis that while retrieval practice is a declarative learning strategy, it shows a similar automatization pattern for item retrieval that is usually observed in skill learning (Squire & Zola, 1996). Logan (1988) quantitatively defined automatization process as a speed up in terms of RTs, which follows a power function and has been accepted as a general description of skill acquisition process (Logan, 1988; Newell & Rosenbloom, 1981). According to this formulation of automatization, the speed-up follows a regular function, characterized by massive gains in the early phases of practice that attenuates with further experience when change of speed reaches an asymptote (Logan, 1988). The function used by Logan (1988) to represent the quantitative properties of automatization is a power function called "the power law of practice," which was described by Newell and Rosenbloom (1981) as:

$$RT = a + bN^{-c}$$

where RT is the time required to do the task, N is the number of blocks, and a, b, and c are unknown parameters (constants). Parameter a (the asymptote of the curve) is the limit of learning determined by the minimum time required to perceive a stimulus and execute a response. Parameter b (the amount to be learned) is the difference between initial and asymptotic performance. Parameter c (learning rate) is the curvature of the fitted curve.

In three experiments we systematically tested the idea that retrieval practice, different from study practice, changes RTs of retrieval in a similar way that was regularly described for automatization during skill acquisition. Moreover, we hypothesized that quantitative properties of automatization during retrieval practice would be associated with the long-term retrieval advantage of tested items over restudied information. In these experiments, we systematically varied the number of practice trials during retrieval/restudy practice and the length of delay between practice and final recall.

Experiment 1

Method

Participants. The participants were 39 Hungarian undergraduate students (native Hungarian speakers; 15 men; age range = 19–29 years, M = 22.4, SD = 2.3). Subjects were recruited at different universities in Budapest, Hungary, and received money for their participation in all experiments. The studies were approved by the Ethical Committee of the Budapest University of Technology and Economics, Hungary. All participants gave written informed consent.

Materials and procedure. The memory task consisted of three phases: an initial learning phase, a practice phase, and a final test phase. Stimuli were 40 Swahili-Hungarian word pairs translated from Nelson and Dunlosky (1994).

In the initial learning phase, participants were presented with all word pairs five times in five consecutive cycles. In each learning cycle, word pairs were presented in random order (5,000 ms/word pair; interstimulus interval [ISI]: 500 ms). Before each learning cycle, participants were instructed to memorize as many word pairs as they could. The relatively high number of initial learning cycles was necessary to achieve a high criterion level which is critical for retrieval practice experiments using no feedback during retrieval practice (see Smith, Roediger, & Karpicke, 2013). The initial learning phase was followed by a 5-min delay while participants were given a list of arithmetic (distractor) tasks.

After the delay, participants practiced the word pairs in six cycles (practice 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 and a retest block (the order of the restudy and retest blocks varied randomly across the learning cycles). We aimed to avoid any confounding effect of using feedback during retrieval practice, therefore participants did not receive feedback (see, e.g., Karpicke et al., 2014). In the restudy blocks, participants saw 20 Swahili words together with their

Hungarian meanings in random order (8,000 ms/word pair; ISI: 500 ms). Before each restudy block, subjects were instructed to memorize the word pairs. In the retest blocks, 20 Swahili words were presented in random order on the computer screen. Participants were instructed to press the Space button on a standard keyboard of the computer when the right answer came to their mind. Participants were allowed to type the Hungarian meanings of the Swahili words only after they pressed the space button. They had a maximum of 8,000 ms to complete one word pair; 8,000 ms after the onset of the stimulus (Swahili word), the next stimulus was presented automatically (preceded by an ISI of 500 ms) independently from the subjects' responses. If subjects answered before the 8,000 ms elapsed, they had to wait until the presentation of the next Swahili word.

Following a 7-day retention interval, participants' memory for all 40 word pairs was tested in the final test phase. Circumstances of the final test were identical to those in the retest condition during the practice phase.

Results

Figure 1a and 1b represent recall rates for the retested items and RTs of correct responses (i.e., time interval between stimulus onset and press of the Space button) in the practice phase of the experiment, respectively. Recall rates as well as RTs were compared between the practice cycles by conducting repeated measures analyses of variance (ANOVAs) with six levels, and then, by conducting simple contrasts with the last (sixth) practice cycle as a reference point.

The ANOVA indicated a significant effect for recall rates, $F(5, 190) = 13.85$, $p < .001$, $\eta_p^2 = .27$, and also for RTs, $F(5, 190) = 37.48$, $p < .001$, $\eta_p^2 = .50$. According to the contrast analysis, recall rate was higher in the last practice cycle than in Cycle 1, $F(1, 38) = 26.16$, $p < .001$, $\eta_p^2 = .41$, and Cycle 2, $F(1, 38) = 15.42$, $p < .001$, $\eta_p^2 = .29$. In addition, RT of correct responses in the last cycle was lower than in all previous practice cycles, Cycle 1: $F(1, 38) = 124.17$, $p < .001$, $\eta_p^2 = .77$; Cycle 2: $F(1, 38) = 80.97$, $p < .001$, $\eta_p^2 = .68$; Cycle 3: $F(1, 38) = 24.77$, $p < .001$, $\eta_p^2 = .40$; Cycle 4: $F(1, 38) = 10.23$, $p < .01$, $\eta_p^2 = .21$; Cycle 5: $F(1, 38) = 8.20$, $p < .01$, $\eta_p^2 = .18$. In brief, whereas recall success increased, RTs decreased during the practice cycles.

In the second cycle of the analysis, as it was described previously, a power curve was fitted to participants' averaged RT data in the practice phase following Logan (1988) to represent the quantitative properties of automatization. An alternative approach would be to use an exponential function and in fact a meta-analytic review of skill learning studies (Heathcote, Brown, & Mewhort, 2000; see also Haider & Frensch, 2002) demonstrated that the exponential function provides a better fit to individual data than the power function. However, in interpreting their results, Heathcote and his colleagues (2000) noted that while there is a consensus among theories of skill acquisition that RTs during practice follow a nonlinear function (e.g., Thorndike, 1913; see also Josephs, Silvera, & Giesler, 1996; Logan, 1988; Newell & Rosenbloom, 1981; Palmeri, 1997), because of the flexible nature of nonlinear functions "it is difficult to determine the exact form of the practice function" (Heathcote et al., 2000, p. 205). We decided to use the power function suggested by Newell and Rosenbloom (1981), because the so-called "power law of practice" became a benchmark test for studies of skill learning (e.g., Logan, 1988, 1992; Palmeri, 1997; see also Haider & Frensch, 2002).

A power curve was fitted to averaged RT data using constrained nonlinear regression (see Cousineau & Lacroix, 2006). Starting values for the unknown parameters were $a = 1$, $b = 0.5$, and $c = 1$. Starting values for the parameters a and b were based on the approximate RT data in the last practice cycle and its approximate difference from data in the first cycle, respectively. The starting value of parameter c was based on the results of Logan (1988). Constrains were as follows: $a \geq 0.25$, $b \geq 0$, and $c \geq 0$. Parameter a (the asymptote) was constrained to not decrease under 0.25 s for psychological plausibility (see Cousineau & Lacroix, 2006). The estimated parameters of the power function were as follows: $a = 0.25$ ($SE = 1.72$), $b = 1.57$ ($SE = 1.71$), and $c = 0.24$ ($SE = 0.32$). The asymptote being 0.25 indicated that the curve is not flattened before reaching the constraint. The measure of goodness of fit (sum of squared errors, *SSE*) showed a high rate of fit, $SSE = 0.004$ (Figure 2).

In the final test, subjects showed superior long-term memory performance, $t(38) = -5.13$, $p < .001$, and lower RTs, $t(38) = 3.10$, $p < .01$, for the retested word pairs than for the restudied items, see Figure 1c and Figure 1d, respectively. To analyze the relationship between performance in the practice cycles (i.e., RTs) and performance on the final test (i.e., recall success), power curves were fitted to individual RT data as well (with the same starting values and constrains as for averaged RT data). We analyzed whether the measure of goodness of fit to the individual power functions was associated with recall accuracy in the final test. The value of *SSE* was negatively correlated with recall rate for the retested word pairs, $r(39) = -.41$, $p < .01$ (Figure 2b; Note that the *SSE* value closer to 0 indicates that the model has a smaller random error component.) The value of *SSE* was correlated with overall recall accuracy as well, $r(39) = -.41$, $p < .01$. However, when we conducted a partial correlation between the *SSE* value and overall recall rate while recall rate for the retested items was used as a covariate, we found no significant correlation between them, $r_{partial}(36) = -.12$, *ns*, indicating that the relationship between the goodness of fit and overall recall accuracy was mediated by recall rate for the retested word pairs.

Discussion

The results of Experiment 1 gave evidence that RTs of retrieval practice followed a power function and the fit to the individual power functions was correlated with long-term retrieval success. To rule out the possibility that this relationship is influenced by the motoric speed up of button press responses during retrieval practice, we carried out a second experiment involving no button press response during practice and varied systematically the number of practice trials each individual item received. The systematic variations of practice numbers for each individual item allowed us to investigate whether study practice also changes the RTs of final recall in a way to fit to a power function.

Experiment 2

Method

Participants. The participants were 62 Hungarian undergraduate students (24 men; age range = 18–28 years, $M = 22.9$, $SD =$

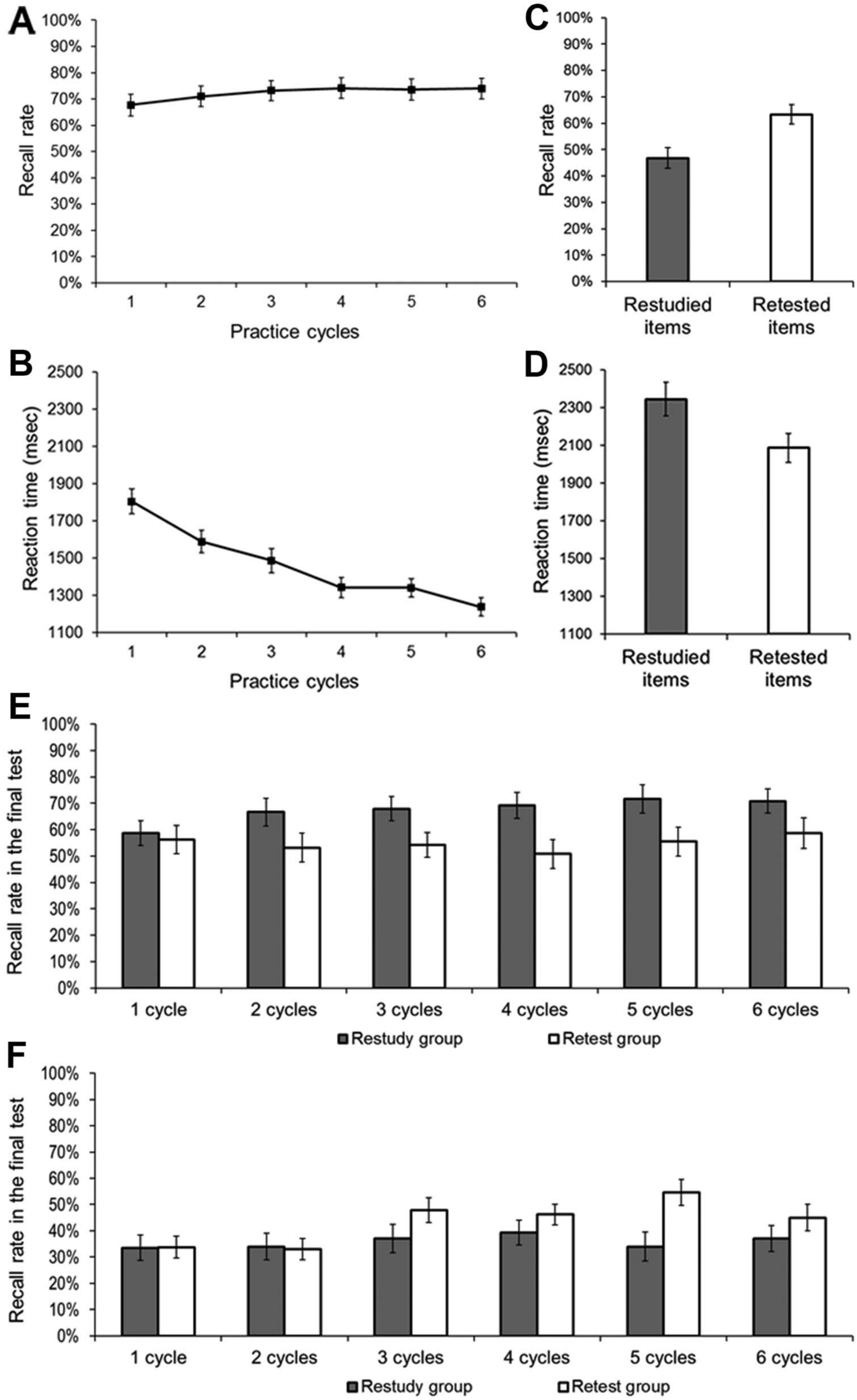

Figure 1. Recall rates and reaction times of correct responses in the practice and final test phases of Experiment 1, 2, and 3. Recall rates (A) and reaction times (RTs; B) for the retested items in the practice phase of Experiment 1. Recall rates (C) and RTs (D) in the final test of Experiment 1. Recall rates in the final test phases of Experiment 2 (E) and Experiment 3 (F). Error bars represent the *SEM*s.

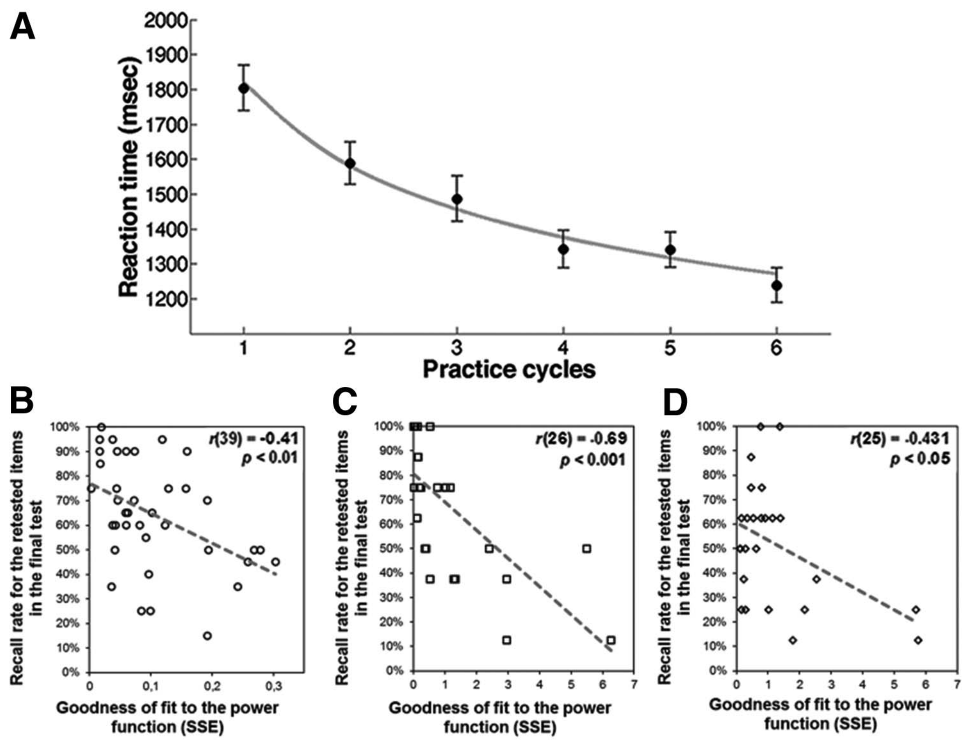

Figure 2. Fitted power function to averaged reaction time data in the practice phase and correlations between the goodness of fit to the individual power functions and final recall performance. Fitted power function to averaged reaction time (RT) data of the practice phase in Experiment 1 (A). Correlation between the goodness of fit to the individual power functions (as indicated by the sum of squared errors of prediction) and recall rate for the retested items in the final test of Experiment 1 (B), Experiment 2 (C), and Experiment 3 (D). Error bars represent the *SEM*s.

2.0). Participants were randomly assigned into either a Restudy or a Retest group. There were 30 participants in the Restudy group (12 men; M_{age} = 23.2 years, SD = 2.0) and 32 participants in the Retest group (12 men; M_{age} = 22.6 years, SD = 1.9).

Materials and procedure. Similar to Experiment 1, the memory task consisted of three phases in Experiment 2: an initial learning phase, a practice phase, and a final test phase. Stimuli were 48 Swahili-Hungarian word pairs. The same 40 word pairs were used as in Experiment 1, and we selected eight additional word pairs from Nelson and Dunlosky (1994).

As in Experiment 1, the initial learning phase consisted of five consecutive learning cycles. Circumstances of the initial learning phase were identical to those in Experiment 1. The initial learning phase was followed by a 5-min delay while participants were given a list of arithmetic (distractor) tasks.

The practice phase consisted of 6 blocks, and the practice blocks contained either 1, 2, 3, 4, 5, or 6 cycles. The word pairs were divided into six parts, and in each block, participants practiced one sixth of the word pairs. In other words, eight word pairs were selected randomly to practice them in one cycle, eight words were selected randomly to practice them in two cycles, and so forth. The number of practice cycles varied randomly between the practice blocks.

Each subject in each practice block practiced the word pairs with one strategy (by repetitive learning in the Restudy group and by repeated retrieval in the Retest group). In the Restudy group, participants saw the Swahili words together with their Hungarian meanings in random order (8,000 ms/word pair; ISI: 500 ms). Before each restudy block, subjects were instructed to memorize the word pairs. In the Retest group, Swahili words were presented in random order on the computer screen. Subjects were not required to press the Space button when the right answer came to their mind (as in Experiment 1), just to type the Hungarian equivalent of the Swahili word. They had a maximum of 8,000 ms to complete one word pair; 8,000 ms after the onset of the Swahili word, the next stimulus was presented automatically (preceded by an ISI of 500 ms) independently from the subjects' responses.

The practice phase was followed by a 15-min delay while subjects were given arithmetic (distractor) tasks. Following the delay, participants' memory for all word pairs was tested in the final test phase. Circumstances of the final test were identical to those in Experiment 1.

Results

For recall rate in the final test phase (Figure 1e), a mixed-design ANOVA was conducted with Cycle (1/2/3/4/5/6) as a within-subjects factor and Strategy (Restudy/Retest) as a between-subjects variable. Whereas the main effect of Cycle, $F(5, 300) = 1.76$, *ns*, $\eta_p^2 = 0.03$, and the Cycle $\times$ Strategy interaction, F(5,

300) = 1.90, *ns*, $\eta_p^2 = 0.03$, were not significant, Strategy had a main effect on recall accuracy, $F(1, 60) = 3.99$, $p < .05$, $\eta_p^2 = 0.06$, indicating that recall rate was higher in the Restudy than in the Retest group.

Because participants with 0% recall rate in any of the six conditions had missing RT data, we conducted a Linear Mixed Model (instead of using an ANOVA) to analyze RTs in the final test phase. The model indicated a significant main effect of Strategy, $F(1, 59.9) = 16.18$, $p < .001$, and a significant Cycle × Strategy interaction, $F(5, 288.4) = 2.97$, $p < .05$, whereas Cycle had no main effect on RT, $F(5, 288.4) = 0.66$, *ns*. These results indicate that when word pairs were practiced in a couple of cycles RT was lower in the Retest than in the Restudy group.

Similar to Experiment 1, power curves were fitted to averaged RT (with the same starting values and constrains as in Experiment 1). In the Restudy group, the estimated parameters were $a = 1.73$ ($SE = 4{,}844.08$), $b = 0.32$ ($SE = 4{,}844.01$), and $c = 0.01$ ($SE = 127.46$). In the Retest group, the estimated parameters were as follows: $a = 0.25$ ($SE = 38.47$), $b = 1.53$ ($SE = 38.41$), and $c = 0.10$ ($SE = 2.65$). The estimated parameter *a* indicated that the asymptote seemed to be reached earlier in the Restudy than in the Retest group. Furthermore, the difference between initial and asymptotic performances (i.e., the amount to be learned) and learning rate were higher in the Retest than in the Restudy condition as indicated by the estimated parameters *b* and *c*, respectively. The *SSE* values were .06 in both groups.

As in Experiment 1, power curves were fitted to individual RT data as well, and we analyzed whether *SSE* values were associated with recall accuracy for those word pairs that were practiced in 6 cycles (either by repetitive learning or repeated retrieval). Significant correlation was found in the Retest group, $r(26) = -.69$, $p < .001$ (Figure 2c), but not in the Restudy condition, $r(28) = -.26$, *ns.* In brief, the goodness of fit to the individual power functions was associated with recall accuracy only for those subjects who practiced by retesting, but not for those participants who practiced by repetitive learning.

Discussion

The results of Experiment 2 showed that the speed up of individual items followed a power function following retrieval practice, even without motoric speed up in the practice phase, and the goodness of fit to the individual power functions was correlated with final recall success. However, in Experiment 2 because of the short-term delay, restudied items showed advantage over retested items, a result is typical in the testing effect literature (e.g., Kornell, Bjork, & Garcia, 2011; Wheeler et al., 2003). Therefore, in a third experiment using the same design as in Experiment 2, we used a long-term delay between practice and final recall to test the hypothesis of the automatization account in the presence of the testing effect.

Experiment 3

Method

Participants. The participants were 58 Hungarian undergraduate students (11 men; age range =19–29 years, $M = 22.2$, $SD = 2.1$). As is Experiment 2, participants were randomly assigned into either a Restudy ($n = 28$; 5 men; $M_{age} = 22.5$ years, $SD = 2.3$) or a Retest group ($n = 30$; 6 men; $M_{age} = 21.9$ years, $SD = 1.9$).

Materials and procedure. In Experiment 3, the same memory paradigm was used as in Experiment 2 with the same parameters and conditions with only one modification: whereas in Experiment 2, there was a 15-min delay between the practice phase and the final test, in Experiment 3, the final test phase was preceded by a 7-day retention interval.

Results

Recall rates and RTs in the final test were analyzed in a similar way as in Experiment 2. A significant main effect of Cycle, $F(5, 280) = 6.40$, $p < .001$, $\eta_p^2 = .10$, and a Cycle × Strategy interaction, $F(5, 280) = 4.27$, $p < .01$, $\eta_p^2 = .07$, were found for recall accuracy (Figure 1f). (Strategy had no main effect on recall rate, $F(1, 56) = 1.72$, *ns*, $\eta_p^2 = .03$.) For RT, a significant Cycle main effect was found, $F[5, 264.8] = 2.53$, $p < .05$. (The Strategy main effect, $F[1, 55.1] = 0.02$, *ns*, and the Cycle × Strategy interaction, $F[5, 264.8] = 1.77$, *ns*, were not significant.)

For averaged RT data, the estimated parameters of the fitted power function were as follows in the two groups: Restudy group – $a = 0.25$ ($SE = 47.03$), $b = 2.23$ ($SE = 46.96$), $c = 0.09$ ($SE = 2.00$), $SSE = 0.06$; Retest group – $a = 2.23$ ($SE = 0.04$), $b = 0.49$ ($SE = 0.09$), $c = 29.83$ ($SE = 0.00$), $SSE = 0.02$. And most important, a similar pattern of relationship emerged between the estimated individual value of *SSE* and recall accuracy: The *SSE* values were correlated with recall accuracy only in the Retest group, $r[25] = -.43$, $p < .05$, but not in the Restudy condition, $r[18] = .09$, *ns* (Figure 2d).

Discussion

The results of Experiment 3 showed that final recall RTs of retested items followed a power function even after a week delay between practice and final recall. The goodness of fit to the individual power functions correlated with final recall success, this time in the presence of the testing effect, because retested items showed a long-term advantage over restudied items. As in Experiment 2, the same correlation between the goodness of fit to individual power functions and long-term recall success did not emerge for the restudied items.

General Discussion

The previous findings gave evidence that the speed up of cued recall following retrieval practice aligned to a power function, which is generally considered as an important quantitative attribute of automatization in skill learning (Logan, 1988; Schneider & Chein, 2003). The measure of goodness of fit to the individual power functions was associated with long-term retention success. Furthermore, the results of Experiment 2 revealed that the speed up of retrieval followed a power function following retrieval practice, even without motoric response in the practice phase, and along with the short-term recall advantage of study practice over retested items. Together these results suggest that retrieval practice decreased the involvement of attentional control and increased the level of automatization of cued recall. Furthermore, because of this automatization process retrieval practice was detrimental for short-term success and beneficial for long-term retention in comparison with study practice.

These results are in line with an influential definition of automatic processes, described as an activation of a sequence of nodes or responses that "nearly always becomes active in response to a particular input configuration" and that "is activated automatically without the necessity for active control or attention by the subject" (Shiffrin & Schneider, 1977, p. 2). It is important that automatic and nonautomatic processes cannot be described with dichotomous properties, such as effortful and effortless (Logan, 1988). Instead, the rate of automatization can be described on a spectrum, and the fit to the practice function is a good indicator of the rate of automatization. We assume that some automatization occurs during study practice as well, however, the rate of automatization is lower than it is observed in the retrieval practice condition and is not associated with long-term retrieval success. The repeated encounter with cue-target pairs during study practice keeps these associations in an active state in short-term memory (STM), which raises the probability of successful retrieval shortly after training without a beneficial effect on long-term retention.

The Relationship Between Automatization and Retention

In an influential article, Schmidt and Bjork (1992) emphasized the similarity between training in motoric and verbal learning. They stressed that those procedures that enhance performance during training can be detrimental in the long term (see also the idea of desirable difficulties in Bjork, 1994). Agreeing with the viewpoint of Schmidt and Bjork (1992), a clear distinction should be made between the indicators of learning (or training) and the indicators of memory retention.

Based on our results, it seems that RT of retrieval during practice predicts higher memory retention, but not successful short-term learning performance. In other words, our results are in line with the suggestion of Schmidt and Bjork (1992) if we assume that retrieval success and retrieval speed are influenced independently by different learning strategies during training. The results of Experiment 2 of the present study raise the possibility that study practice increases the short-term activation and accessibility of the learned information without changing the speed of retrieval, whereas retrieval practice speeds up retrieval without raising effectively the short-term accessibility of learned associations. Therefore, we suggest placing more emphasis on retrieval speed during training as an indicator of long-term benefit of training.

Bjork (2011) argued convincingly that forgetting is an important facilitator of retention, experimental manipulations that decrease performance during training (such as spacing study trials or changing context during encoding trials) enhance retention. Agreeing with this idea, here we suggest that there are important mediator factors in this relationship, namely, the involvement of attentional control and the level of automatization of cue-target reactivation during training. Accordingly, it seems that retrieval practice reduces the involvement of control processes in retrieval, as it is revealed by RT analyses. Therefore, we suggest that the nonlinear speed up of retrieval is an indicator of storage strength and retention, whereas short-term retrieval success is an indicator of retrieval strength (see Bjork & Bjork, 1992). Along with this assumption, recently we found that retrieval practice trials decreased mental effort and the involvement of attentional control during retrieval as measured by pupil dilation. In contrast, this relationship was not detected during study practice, suggesting that this latter type of learning strategy keeps control processes highly involved in learning during training (Racsmány, Pajkossy, & Szőllősi, 2017). This assumption is also compatible with the results of Mulligan and Picklesimer (2016) who found that retrieval-practiced information were less sensitive than restudied items for divided attention situations where control processes are heavily loaded by a secondary task.

Although we suggest that the level of automatization is an important factor in understanding the long-term benefit of retrieval practice, we are aware that the concept of automaticity is without unanimous understanding. Because providing an in-depth analysis of the concept and the features of automaticity is far beyond the scope of the present article, we shortly summarize only those aspects of automaticity and skill learning that seems to be relevant for our account.

Approaches of automatization differ in whether automatic and controlled processes are viewed as all-or-none concepts or as being on a spectrum (see Moors & De Houwer, 2006 for a review). There is a tentative consensus on the role of attentional control as a central distinguishing feature of automatic and nonautomatic processes (but see Shiffrin & Schneider, 1977, for the distinction of two kinds of attention allocation). That is, automatic processes require minimal attentional resources in contrast with the high attentional requirements of the nonautomatic tasks (see Hasher & Zacks, 1979). Logan (1985) described automatization process as a shift from algorithm computation to single-step memory retrieval. In contrast, Anderson (1992) proposed that reduction of procedural steps and strengthening of algorithms are the background mechanisms of automatization. Anderson and colleagues found the same power-law speed-up with sentence recognition task and fact retrieval as it is usually observed in skill learning (Pirolli & Anderson, 1985).

Our data and the scope of our experiments are not meant to contrast these theories. Instead, our starting point in planning our experiments was the view that suggests that automaticity could be investigated through the presence of certain features of automaticity (Moors & De Houwer, 2006). According to this approach, all features of automaticity have indicators by which they could be operationalized. According to Moors and De Houwer (2006), there are main features of automatic processes, namely, they are unintentional, uncontrolled, goal-independent, autonomous, purely stimulus driven, long-term efficient, and fast. In our study, we focused on long-term efficiency and fast processing using RT speed up and long-term retrieval success as operationalized measures. It should be the aim of future studies to investigate the relationship between the remaining features of automaticity and retrieval-based learning. As it was argued by Logan (1985) "Automaticity and skills are closely related but are not identical" (Logan, 1985, p. 367), it is better to conceptualize automaticity as to be a component of skills. The features of automaticity are characteristic to complex skills, which are also characterized with reduced attention, multitask tolerance, and reduced cognitive effort (Christensen, Sutton, & McIlwain, 2016). As it was suggested previously, retrieval-based learning is also characterized with these features; however, as is the case with complex skills, automaticity is only a component of retrieval-based learning and the two are not identical.

Automatization and the Theoretical Accounts of the Testing Effect

Here we suggest that retrieval practice is advantageous for long-term learning through building up direct cue-target associations, which will ensure that retrieval cues will effectively elicit the targeted responses without involving algorithmic searching processes (Logan, 1988). Framing the testing effect and retrieval-based learning in general, as they are influenced by automatization of cue-target assembling, fits well with a range of experimental findings and contemporary models of human retrieval (Keresztes et al., 2014; Logan, 1988; Raichle et al., 1994; Roediger & Butler, 2011; Schneider & Chein, 2003). Our account is also in line with the result of a recent neuroimaging study showing that the activation of the dorsolateral prefrontal cortex, one of the central components of the attentional control network, monotonically decreased as a function of repeated successful retrieval for items subsequently remembered, but not for memories subsequently forgotten (Karlsson Wirebring et al., 2015).

We think that our account concerning the role of automatization in retrieval-based learning is compatible with some of the contemporary theories of the testing effect. For instance, the idea of transfer-appropriate processing proposed that retrieval performance is higher when cognitive processes recruited during the original learning or training phases and the final test are overlapping (Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977). Although this approach does not specify the underlying processes of the testing effect, it is compatible with the automatization account that assumes that retrieval practice strengthens the direct cue-target associations and decreases the involvement of attentional control both during practice and final recall. As a consequence, similar cue-target reactivation processes will take place during training and final recall without the effects of the actual level of the control system (Logan, 1988; Moors & De Houwer, 2006).

The automatization account is also compatible with some aspects of the episodic context account of retrieval-based learning (Karpicke et al., 2014). This theory proposes that during retrieval practice, participants use available cues to reconstruct learned information. This reconstruction process includes the reinstatement and updating of the temporal context of the learning phase; as a consequence, repeatedly retrieved items become associated with many temporal contexts. The most important consequence of this rich contextual information is the restriction of the search set. In other words, retrieval practice decreases the number of subsets of items which can be potentially associated to retrieval cues, consequently the retrieval cues uniquely specify the target items (Karpicke et al., 2014). This idea is completely compatible with the automatization account, if we assume that rich contextual information during the final test restricts the search set through automatized and fast cue-target reactivation without the involvement of the interference-sensitive control system.

We are aware of the limitations of the present study and our account of the testing effect. Here we used only cued recall both during practice and final recall, and it will be important to test the predictions of the automatization account by conducting experiments using free recall or recognition tasks and by using paradigms where the test formats of the practice and final test phases are not identical. It is also an open question whether our account is suitable for explaining the transfer effect of retrieval practice, as it was shown that retrieval practice enhanced transfer to a new knowledge domain compared with study practice (Butler, 2010). Although we did not investigate the issue of transfer and its relationship to automatization, there are studies that found causal relationship between the features of automatization during training and the transfer to similar problems in simple algebra transformation problems (Cooper & Sweller, 1987) and computer programming (Van Merrienboer & Paas, 1990).

Altogether, the previous findings point out that retrieval practice leads to a diminishing involvement of attentional control in declarative retrieval and preserves long-term knowledge through fast and automatized processing of specific cue-target associations. Memories practiced through retrieval show low-level of forgetting and are resistant to injuries and interference effects (Karpicke et al., 2014), just like skills, a similarity raising the possibility to use retrieval-based learning techniques for populations with seriously impaired declarative learning functions. The automatization account of the testing effect also lends itself for future hypothesis testing by assuming that retrieval practiced items show the characteristic features of automatization described above.

References

Anderson, J. R. (1992). Automaticity and the ACT* theory. *The American Journal of Psychology, 105,* 165–180. http://dx.doi.org/10.2307/1423026

Atkinson, R. L. (1996). *Hilgard's introduction to psychology* (Vol. 12). Philadelphia, PA: Harcourt Brace College Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (2011). On the symbiosis of learning, remembering, and forgetting. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 1–22). London, England: Psychology Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

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

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. http://dx.doi.org/10.1037/a0017021

Christensen, W., Sutton, J., & McIlwain, D. J. F. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language, 31,* 37–66. http://dx.doi.org/10.1111/mila.12094

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79,* 347–362. http://dx.doi.org/10.1037/0022-0663.79.4.347

Cousineau, D., & Lacroix, G. L. (2006). Getting parameters from learning data. *Tutorials in Quantitative Methods for Psychology, 2,* 77–83. http://dx.doi.org/10.20982/tqmp.02.2.p077

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 392–400. http://dx.doi.org/10.1037//0278-7393.28.2.392

Hanawalt, N. G. (1937). Memory trace for figures in recall and recognition. *Archives de Psychologie, 31,* 216–223.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388. http://dx.doi.org/10.1037/0096-3445.108.3.356

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7,* 185–207. http://dx.doi.org/10.3758/BF03212979

Hikosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology, 12,* 217–222. http://dx.doi.org/10.1016/S0959-4388(02)00307-0

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior, 17,* 649–667. http://dx.doi.org/10.1016/S0022-5371(78)90393-6

Josephs, R. A., Silvera, D. H., & Giesler, R. B. (1996). The learning curve as a metacognitive tool. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 510–524. http://dx.doi.org/10.1037/0278-7393.22.2.510

Karlsson Wirebring, L., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *The Journal of Neuroscience, 35,* 9595–9602. http://dx.doi.org/10.1523/JNEUROSCI.3550-14.2015

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 61, pp. 237–284). New York, NY: Elsevier.

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. http://dx.doi.org/10.1093/cercor/bht158

Keresztes, A., & Racsmány, M. (2013). Interference resolution in retrieval-induced forgetting: Behavioral evidence for a nonmonotonic relationship between interference and forgetting. *Memory & Cognition, 41,* 511–518. http://dx.doi.org/10.3758/s13421-012-0276-3

Kolers, P. A., & Roediger, H. L., III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23,* 425–449. http://dx.doi.org/10.1016/S0022-5371(84)90282-2

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. http://dx.doi.org/10.1016/j.jml.2011.04.002

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience, 10,* 908–914. http://dx.doi.org/10.1038/nn1918

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59,* 277–287. http://dx.doi.org/10.1080/02701367.1988.10609373

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1787–1794. http://dx.doi.org/10.1037/xlm0000012

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367–386. http://dx.doi.org/10.1037/h0080066

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

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105,* 317–339.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297–326. http://dx.doi.org/10.1037/0033-2909.132.2.297

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

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

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

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354. http://dx.doi.org/10.1037/0278-7393.23.2.324

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136–153. http://dx.doi.org/10.1037/0278-7393.11.1.136

Racsmány, M., & Keresztes, A. (2015). Initial retrieval shields against retrieval-induced forgetting. *Frontiers in Psychology, 6,* 657. http://dx.doi.org/10.3389/fpsyg.2015.00657

Racsmány, M., Pajkossy, P., & Szőllősi, Á. (2017). *Retrieval practice decreases mental effort and attentional control of recall: Evidence revealed by pupillometry*. Manuscript under review.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4,* 8–26. http://dx.doi.org/10.1093/cercor/4.1.8

Ratcliff, R., & McKoon, G. (2000). Memory models. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 571–582). New York, NY: Oxford University Press.

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. http://dx.doi.org/10.1016/j.tics.2010.09.003

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

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20,* 1053–1058. http://dx.doi.org/10.1111/j.1467-9280.2009.02415.x

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00029.x

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science, 27,* 525–559. http://dx.doi.org/10.1207/s15516709cog2703_8

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190. http://dx.doi.org/10.1037/0033-295X.84.2.127

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. http://dx.doi.org/10.1037/a0033569

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 13515–13522. http://dx.doi.org/10.1073/pnas.93.24.13515

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. http://dx.doi.org/10.1037/a0013082

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221. http://dx.doi.org/10.1037/0278-7393.4.3.210

Thorndike, E. L. (1913). *Educational psychology: The psychology of learning* (Vol. 2). New York, NY: Teachers College Press. http://dx.doi.org/10.1037/13051-000

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. http://dx.doi.org/10.1016/j.neuroimage.2013.03.071

Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior, 6,* 273–289. http://dx.doi.org/10.1016/0747-5632(90)90023-A

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571–580. http://dx.doi.org/10.1080/09658210244000414

Whitten, W. B., II, & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16,* 465–478. http://dx.doi.org/10.1016/S0022-5371(77)80040-6

Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 677–691. http://dx.doi.org/10.1037/0278-7393.16.4.677

Received January 18, 2017
Revision received April 4, 2017
Accepted April 4, 2017 ■