

Testing Enhances Subsequent Learning in Older Adults

Bernhard Pastötter
University of Trier

Karl-Heinz T. Bäuml
Regensburg University

Interference susceptibility has been suggested to be a major factor for episodic memory impairment in healthy older adults. Previous work has shown that retrieval practice can reduce proactive interference and thus enhance learning and memory in younger adults, a finding referred to as the forward effect of testing in the literature. This study examined the late developmental trajectory of the forward effect in middle-aged and older adults (40 to 79 years of age). Participants studied three lists of items in anticipation of a final cumulative recall test. In the testing condition, participants were tested immediately on Lists 1 and 2 after initial study, whereas in the restudy condition, they restudied Lists 1 and 2. In both conditions, participants were tested immediately on List 3. The results of the immediate List-3 recall test showed a reliable forward effect of testing, with interim testing of Lists 1 and 2 enhancing List-3 recall and reducing the number of prior-list intrusions. Notably, this effect of testing was found independent of participants' age. These results suggest that retrieval practice can reduce proactive interference in middle-aged and older adults. Together with recent findings on the presence of the backward effect of testing in older adults—that is, improved recall of studied material after retrieval versus restudy, these findings indicate that retrieval practice can be a very powerful tool to delimit memory impairment in older age.

Keywords: cognitive aging, episodic memory, retrieval practice, forward testing effect

Older adults often show impaired episodic memory compared with younger adults (for a review, see Park & Festini, 2017). One explanation for this impairment is that older adults are more susceptible to memory interference than younger adults (Biss, Campbell, & Hasher, 2013; Hasher & Zacks, 1988; Jacoby, Debnar, & Hay, 2001). Interference susceptibility can, for instance, be examined in proactive interference (PI) tasks, in which individuals' memory for recently studied target information is commonly impaired by the previous study of other nontarget information (Underwood, 1957). PI has been attributed to both encoding and retrieval factors. Temporal discrimination theory is the most prominent retrieval account of PI. It assumes that the additional study of nontarget material causes a temporal discrimination problem between target and nontarget information at test, reducing memory accessibility of the target material (Crowder, 1976; Wixted & Rohrer, 1993). In addition, PI has been attributed to impaired encoding. According to the integrated encoding–retrieval account, for instance, attentional resources may decrease and memory load

increase from initial encoding of the nontarget information to the final encoding of the target information, which can attenuate encoding of the target items (Kliegl, Pastötter, & Bäuml, 2015).

Following these accounts, enhanced interference susceptibility in older adults may arise from reduced attentional resources and increased memory load during encoding and/or impaired temporal discrimination ability during retrieval. This view is consistent with the inhibitory deficit account, which assumes that aging is associated with a decreased ability to ignore nontarget information and focus attention on the target information (Hasher & Zacks, 1988), leaving individuals with an enhanced interference level and inducing recall impairment. From a practical point of view, it is therefore important to identify factors that can diminish the interference susceptibility of older adults. One such factor may be retrieval practice, which has been shown to reduce PI in younger adults.

Retrieval practice can have a number of beneficial effects on long-term memory and learning. For instance, a very prominent benefit of retrieval practice is the (backward) testing effect (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006). This effect refers to the finding that when participants repeatedly retrieve versus restudy previously studied item material, the retrieved items are better recalled on a delayed memory test than the restudied items. The testing effect generalizes across a wide range of materials and experimental settings (for reviews, see Karpicke, 2017; Roediger & Butler, 2011). Although the effect has mostly been examined in younger adults, recent research has also established the effect in children (e.g., Fritz, Morris, Nolan, & Singleton, 2007; Lipowski, Pyc, Dunlosky, & Rawson, 2014) and older adults (e.g., Coane, 2013; Meyer & Logan, 2013). It arises in both healthy and clinical populations, including patients with multiple

This article was published Online First October 18, 2018.

Bernhard Pastötter, Department of Psychology, University of Trier; Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University.

The authors thank J. Rupprecht for her help with data collection. The dataset is available on request; the raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

Correspondence concerning this article should be addressed to Bernhard Pastötter, Department of Psychology, University of Trier, 54286 Trier, Germany. E-mail: pastoetter@uni-trier.de

sclerosis, traumatic brain injury (TBI), and Alzheimer's disease (Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Sumowski, Chiaravalloti, & DeLuca, 2010; Sumowski, Wood, et al., 2010). The effect has been attributed to quite different mechanisms, including the view that retrieval practice, more than restudy, induces elaborative or deep processing of the practiced information (Carpenter, 2009), and the view that retrieval practice, more than restudy, enhances contextual processing of the practiced information (Karpicke, Lehman, & Aue, 2014).

Another prominent benefit of retrieval practice is the forward effect of testing, which refers to the finding that retrieval practice of previously studied information can increase retention of subsequently studied *other* information (e.g., Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Szpunar, McDermott, & Roediger, 2008). The forward effect is striking because it is an effect on the learning of new information not necessarily related to the retrieved information. The effect can be studied in interim-testing tasks. In this type of task, participants study several (e.g., three) lists of items in anticipation of a final cumulative recall test. Between study of each pair of lists (i.e., Lists 1 and 2), participants restudy the words from the most recent list (either List 1 or List 2), or are engaged in retrieval practice on these items. After study of all lists, participants are asked to recall the words from the final critical list (i.e., List 3). The typical finding is that testing of the preceding lists (Lists 1 and 2) enhances recall of the critical list (List 3) and reduces the number of prior-list intrusions in the immediate List-3 recall test.

The forward effect is typically attributed to PI reduction. The idea is that recall testing between the study of lists promotes contextual list segregation, which may enhance temporal differentiation between preceding and critical lists at test (Bäuml & Kliegl, 2013; Divis & Benjamin, 2014; Szpunar et al., 2008) and may also increase attentional resources during critical list encoding (Pastötter et al., 2011; Pastötter, Engel, & Frings, 2018). The forward effect is a robust phenomenon that has been replicated in a number of studies employing different item materials and different experimental settings (for reviews, see Pastötter & Bäuml, 2014; Yang, Potts, & Shanks, 2018). It has been reported in young adults as well as in middle adulthood, in both healthy people and clinical patients with severe TBI (around 40 years of age; Pastötter, Weber, & Bäuml, 2013). Regarding the early developmental trajectory of the effect, Aslan and Bäuml (2016) observed that the effect develops relatively late in childhood. These researchers reported a reliable forward effect in children around 9 years of age but no such effect in children around 7 years of age. To the best of our knowledge, no study has yet examined the late developmental trajectory of the forward effect from middle adulthood to older age. In the present study, we aimed at filling this gap.

What might be expected regarding the late developmental trajectory of the forward effect of testing from middle adulthood to older age? Based on the finding that the effect develops relatively late in childhood (Aslan & Bäuml, 2016) and the idea that cognitive effects sometimes follow a (reversed) U-shaped developmental function (see Pauls, Macha, & Petermann, 2013), the expectation may arise that the forward effect decreases from middle adulthood to older age. Like the young children, the older adults may thus show an inability to combat PI. However, there is evidence from other experimental tasks, like the list-method directed forgetting (LMDF) task, that PI reduction can be efficient in

older adults. In the LMDF task, participants study two lists of items and, after study of the first list, they receive a cue either to forget or to continue remembering this list. After study of the second list, a memory test for both lists is conducted in which all participants are asked to recall the items of the two lists, regardless of original cuing. The typical finding is that the forget cue impairs recall of the first list but enhances recall of the second list, with the enhancement being attributed to PI reduction (for reviews, see Pastötter, Tempel, & Bäuml, 2017; Sahakyan, Delaney, Foster, & Abushanab, 2013). It is important to note that the enhancement effect in LMDF has been found to be equally present in younger and older adults (Aslan & Bäuml, 2013; Zellner & Bäuml, 2006), indicating efficient PI reduction in older adults in this experimental task. Critically, if older adults' ability to reduce PI is not restricted to the LMDF task, but generalizes to other multilist learning tasks, then the forward effect of testing may not decrease with age, but be similar in size from middle adulthood to older age.

The present study addressed the issue as we examined the late developmental trajectory of the forward testing effect, employing four age-decade groups (i.e., 40s, 50s, 60s, and 70s). Participants studied three lists of items (Lists 1–3), which they were asked to remember for a final cumulative recall test. In the testing condition, participants were tested immediately on Lists 1 and 2 after studying each single list, whereas in the restudy condition, they restudied Lists 1 and 2 directly upon a list's study. In both conditions, participants were tested immediately on List 3 and later took part in the final cumulative test. With regard to the forward effect, the critical variables were correct recall and the number of prior-list intrusions in the immediate List-3 recall test. We expected to replicate the previous finding of a reliable forward effect in adults in the 40s group (Pastötter et al., 2013). Regarding the developmental trajectory of the forward effect, two possible expectations arise. The one expectation follows Aslan and Bäuml's (2016) recent finding of no forward effect of testing in young children and the idea of a (reversed) U-shaped developmental function (Pauls et al., 2013), suggesting that the effect may decrease with older age and, at a high age level, be even absent. The other, alternative expectation is that the effect is equally present in all four age-decade groups. This expectation is based on evidence from LMDF work that suggests efficient PI reduction in older adults (Aslan & Bäuml, 2013; Zellner & Bäuml, 2006). Indeed, if older adults' ability to reduce PI is not restricted to the LMDF task but generalized to the present interim-testing task, then no influence of participants' age on the forward effect of testing should be observed.

Method

Participants

Sample size was calculated a priori using G*Power (Version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007). Required sample size was calculated for the interaction effect in a 2×4 analysis of variance (ANOVA) with fixed effects and the two between-subjects factors, experimental practice condition (testing, restudy) and age-decade group (40s, 50s, 60s, 70s). The following input parameters were chosen. Type-I error probability was set to .05, power was set to .90, and size of the interaction effect was set to medium, $f = .25$. In so doing, we followed the developmental study by Aslan and Bäuml (2016), which showed a medium

Condition \times Age interaction effect for the forward effect of testing in early childhood. Based on these input variables, G*Power suggested a total sample size of 231.

Participants were 240 volunteers recruited from the community. Four age-decade groups were considered, with 60 participants in each group. Participants were 40 to 49 years of age in the 40s group, 50 to 59 years in the 50s group, 60 to 69 years in the 60s group, and 70 to 79 years in the 70s group (see Table 1 for descriptives). In each age-decade group, participants were randomly assigned to one of the two experimental practice conditions. A two-way ANOVA on age with practice condition and age-decade group showed no significant main effect of practice condition and no significant interaction between the two variables, $F_s < 1.5$, $p_s \geq .224$. All participants were tested individually. The study was approved by the ethical review committee at Regensburg University Medical Center, Regensburg, Germany. Informed consent was obtained from all participants.

Additional two-way ANOVAs with practice condition and age-decade group were calculated for the following variables: participants' years of formal education; participants' performance on the backward-counting task, the Working Memory subscale of the *Wechsler Adult Intelligence Scale* (4th ed.; WAIS-IV; Wechsler, 2008), which is a short-term/working-memory task that estimates fluid intelligence; and participants' performance on the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; [Multiple Choice Vocabulary Intelligence Test]; Lehrl, 2005), which is a German vocabulary test that estimates crystallized intelligence (see Table 1 for descriptives). The analyses showed no significant main effect of practice condition, $F_s \leq 2.82$, $p_s \geq .094$, and no significant interaction between the factors of practice condition and age-decade group, $F_s < 1$, $p_s \geq .599$. Significant main effects of age-decade group were found for participants' years of formal education, $F(3, 232) = 7.82$, $MSE = 14.92$, $p < .001$, $\eta_p^2 = .091$, and participants' performance on the backward-counting task, $F(3, 232) = 8.80$, $MSE = 4.63$, $p < .001$, $\eta_p^2 = .102$. Indeed, participants in the 70s group had significantly shorter formal education than participants in the three other groups, $p_{\text{Tukey}} \leq .005$. In the backward-counting task, the 40s group performed better than the 70s, and the 50s group performed better than the 60s and 70s groups, $p_{\text{Tukey}} \leq .019$. Age-decade groups did not differ in their performance on the MWT-B, $F(3, 232) = 1.83$, $MSE = 15.71$, $p = .143$.

Materials

Three study lists (A, B, C) were constructed, each containing 10 unrelated concrete German nouns drawn from the CELEX database (Duyck, Desmet, Verbeke, & Brysbaert, 2004). The nouns were of medium frequency (mean frequency per million = 20.6 ($SD = 32.3$)) and were four to eight letters long.¹ The assignment of study lists A, B, and C to experimental Lists 1, 2, and 3 was counterbalanced between practice conditions and age-decade groups. Number of lists (three) and number of items per list (10) were chosen to be the same as in the clinical research study by Pastötter et al. (2013).

Design

The experiment had a between-participants 2 (practice condition) \times 4 (age-decade group) design. In the testing condition, Lists 1 and 2 were immediately tested after initial study; in the restudy condition, Lists 1 and 2 were restudied after initial study.

Procedure

Participants were informed that they would be presented with three lists of items. They were encouraged to pay close attention to the presented items for a final cumulative free recall test at the end of the experiment, in which all of the previously presented items would be tested. Participants were further told to expect various activities that may follow the presentation of each single list: backward counting in steps of two, restudy of items from a list that had just been studied in a new random order, and immediate free-recall testing of items that had just been studied. We pretended that activities following each list were determined randomly. In fact, however, activities following Lists 1 and 2 were always the same and depended on the experimental practice condition (see Figure 1). It was highlighted that, regardless of whether or not a list's items were tested immediately, all lists' items would be tested on the final cumulative recall test (for a similar procedure, see Szpunar et al., 2008, or Pastötter et al., 2011).

The experimenter presented the three 10-item lists both visually and orally, successively on 2.9×4.1 -inch (7.4×10.4 -cm) index cards, with a random-order presentation rate of 6 s. The experimenter dealt the index cards, consecutively placing them on top of each other and putting the deck away after presenting the last item of a list. Font size was 50 points, font color was black on white background. During presentation, each item was read out loud by the experimenter. Subsequent to the presentation of each list (1 min), all participants counted backward in steps of two from a three-digit number (smaller than 400) for 30 sec. Experimental practice conditions differed in participants' activity, which followed backward counting after Lists 1 and 2. In the testing condition, participants were given 1 min to tell the experimenter in any order they wished as many items as they could remember from the just-studied list (immediate recall of Lists 1 and 2). In the restudy condition, participants were re-presented the just-studied lists in new random order (1 min each). In both conditions, backward counting after List 3 was followed by an immediate List-3 free-recall test; participants were given 1 min to tell the experimenter in any order they wished as many items they could remember from List 3. All responses were given orally by the participants and were written down by the experimenter.

After the immediate List-3 recall test, participants did an easy mathematical distractor task for 3 min, in which they added pairs of one- and two-digit numbers as fast and correctly as possible. Finally, participants took part in the cumulative recall test, in which they were asked to separately recall the three

¹ The CELEX database does not involve arousal ratings of the German words. Post-hoc analysis of arousal levels of the 24 (out of 30) words listed in the Berlin Affective Word List-Reloaded (BAWL-R; Võ et al., 2009) revealed a mean arousal level of 2.43, which is below average (2.76) in the BAWL-R.

Table 1
Demographic and Assessment Measures as a Function of Age-Decade Group and Practice Condition

Independent variables		Measures			
Age-decade group	Condition	Age	Years of education	Backward counting	MWT-B
40s	Testing	45.10 (3.02)	13.75 (3.36)	8.53 (1.91)	29.37 (3.34)
	Restudy	46.17 (3.04)	15.00 (3.24)	8.67 (2.23)	30.13 (3.04)
50s	Testing	53.70 (2.72)	14.97 (3.75)	9.17 (2.48)	31.33 (2.92)
	Restudy	53.40 (2.97)	15.55 (3.35)	9.20 (2.27)	30.80 (3.15)
60s	Testing	63.43 (2.86)	14.20 (5.47)	8.26 (2.33)	30.87 (4.52)
	Restudy	63.57 (2.73)	14.57 (4.15)	7.80 (2.07)	30.73 (3.60)
70s	Testing	73.97 (3.17)	11.43 (3.36)	7.23 (1.55)	29.13 (6.08)
	Restudy	74.93 (3.17)	12.58 (3.71)	7.27 (2.24)	30.37 (4.07)

Note. MWT-B = Multiple Choice Vocabulary Intelligence Test; means and standard deviations (in brackets).

lists' items. Recall time for each list was 1 min and the experimenter noted participants' verbal responses. Recall order in the cumulative recall test was counterbalanced for Lists 1 and 3, which were tested either first or second. List 2 was always tested last. Note that final List-1 recall provided a rough measure of the backward-testing effect (with retroactive interference from Lists 2 and 3 and, when List 1 was recalled last, prior recall of List 3 on the final test) and List-3 recall provided a rough measure of the forward testing effect (with prior immediate recall of List 3 and prior recall of List 1 when List 3 was recalled last). In contrast, final List 2 recall provided a mixed measure of backward and forward effects (with prior recall of Lists 1 and 3 on the final recall test). To anticipate the results, final list recall order did not influence any of the interactions with the factors of practice condition and age-decade group (all $ps \geq .245$). The whole experiment lasted approximately 15 min. After the experiment, the WAIS-IV (Wechsler, 2008) backward-counting task and the MWT-B (Lehrl, 2005) were administered, which took about 10 min. Upon completion of these tasks, participants were thanked and debriefed.

Data Analysis

Both correct recall and number of intrusions were examined. Correct recall was examined in all three lists' immediate and final recall tests. Number of intrusions was examined in the immediate List-3 recall test; List 1 and List 2 items that were falsely recalled by participants in the immediate List-3 recall were considered as intrusions. Both classical and Bayesian ANOVAs with the testing/restudy and age-decade group variables were calculated (for Bayesian analysis, JASP 0.8.5.1 software with default settings was used; see Wagenmakers et al., 2018). Post hoc analyses were conducted employing Tukey's honestly significant difference testing.

Results

Immediate Recall: List 3

Regarding the forward effect of testing on immediate List-3 recall (see Figure 2A and Table 2), a two-way ANOVA revealed a main effect of practice condition, $F(1, 232) = 24.70$, $MSE = 487.59$, $p < .001$, $\eta_p^2 = .096$, and a main effect of age-decade

group, $F(3, 232) = 18.34$, $MSE = 487.59$, $p < .001$, $\eta_p^2 = .192$, but no interaction between the two variables, $F(3, 232) < 1$. Indeed, participants recalled more items in the testing condition than in the restudy condition (56.6% vs. 42.4%; see Table 2 for pairwise comparisons in each age-decade group). In addition, the 40s (64.0%) and 50s (54.8%) groups recalled more items than the 60s (42.0%) and 70s (37.2%) groups, $ps_{\text{Tukey}} \leq .009$.² Consistent with the nonfinding of a significant interaction, Bayesian analysis showed that the data were 15.16 times more likely under the two main-effects model than under the model that added the interaction, thus providing strong evidence against including the interaction term in the model. Together, these results suggest that the forward effect of testing on immediate List-3 recall was reliable and equally present in all four age-decade groups.

Regarding prior-list intrusions (see Figure 2B and Table 2), a two-way ANOVA showed a main effect of practice condition, $F(1, 232) = 7.84$, $MSE = 1.12$, $p = .006$, $\eta_p^2 = .033$, and a main effect of age-decade group, $F(3, 232) = 2.83$, $MSE = 1.12$, $p = .039$, $\eta_p^2 = .035$, but no interaction between them, $F(3, 232) < 1$. Indeed, participants showed fewer intrusions in the testing condition than in the restudy condition (0.55 vs. 0.93; see Table 2 for pairwise comparisons in each age-decade group). The 40s group (0.43) showed fewer intrusions than the 70s group (0.98), $p_{\text{Tukey}} = .025$, 50s = 0.73, 60s = 0.82. Consistently, Bayesian analysis revealed that the data were 16.53 times more likely under the two-main-

² Note that the results did not change when controlling for differences in participants' years of formal education, performance on the backward-counting task, and performance on the MWT-B. Indeed, an analysis of covariance (ANCOVA) with the two factors of practice condition and age-decade group and the three covariates of participants' years, formal education, performance on the backward-counting task, and performance on the MWT-B, revealed a main effect of practice condition, $F(1, 229) = 29.39$, $MSE = 461.18$, $p < .001$, $\eta_p^2 = .114$, and a main effect of age-decade group, $F(3, 229) = 14.42$, $MSE = 461.18$, $p < .001$, $\eta_p^2 = .159$, but no interaction between the two factors, $F(3, 229) < 1$. Participants' years of formal education showed a positive influence on immediate List-3 recall, $F(1, 229) = 6.23$, $MSE = 461.18$, $p = .013$, $\eta_p^2 = .026$, whereas performances on the backward-counting task and the MWT-B had no influence on immediate List-3 recall, $F_s(1, 229) \leq 1.41$, $ps \geq .236$. The correlation between participants' years of formal education and immediate List-3 recall, with age partialled out, was $r(237) = .163$, $p = .012$.

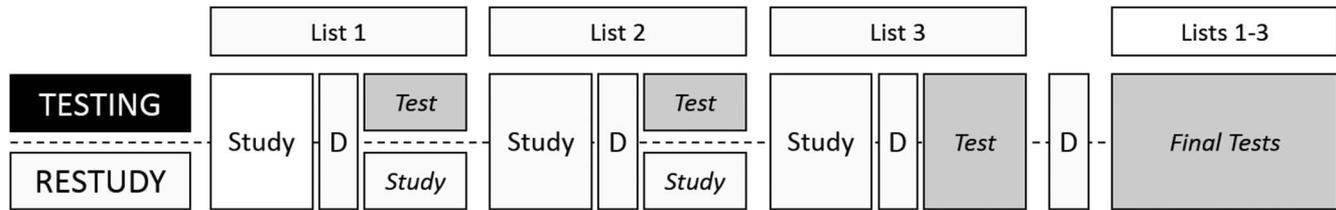


Figure 1. Procedure of the memory task. In both the testing and the restudy condition, participants studied three lists of items. Each list was followed by a short distractor (D) task. List 3 was tested immediately in both conditions; Lists 1 and 2 were tested immediately after the D task in the testing condition, but restudied in the restudy condition. All three lists were tested in the final cumulative recall test.

effects model than under the model that added the interaction.³ Thus, testing of Lists 1 and 2 reduced the number of prior-list intrusions on the immediate List-3-recall test, independent of age-decade group.

Lists 1 to 3 in the Testing Condition

Naturally, immediate recall of Lists 1 to 3 was measured in the testing condition only (see Table 2). A two-way ANOVA with the within-subjects variable, list (List 1 vs. List 2 vs. List 3), and the between-subjects variable, age-decade group, revealed a main effect of list, $F(2, 232) = 20.26$, $MSE = 194.28$, $p < .001$, $\eta_p^2 = .149$, and a main effect of age-decade group, $F(3, 116) = 9.22$, $MSE = 791.95$, $p < .001$, $\eta_p^2 = .193$, but no interaction between the variables, $F(6, 232) < 1$. Indeed, List-1 recall (65.9%) was higher than both List-2 (55.5%) and List-3 recall (56.6%), $ps < .001$, whereas there was no difference between List-2 and List-3 recall, $p = .557$, Holm-corrected. The 40s group (70.3%) recalled more items than the 60s (54.6%) and 70s (49.9%) groups, $ps_{\text{tukey}} \leq .002$; in addition, the 50s group (62.6%) recalled more items than the 70s group, $p_{\text{tukey}} \leq .002$. Bayesian analysis revealed that the two-main-effects model was the best model to explain the data. Indeed, the data were 30.68 times more likely under the two-main-effects model than under the second best model, which added the interaction.

Final Recall

In the first step, an overall three-way ANOVA of list, practice condition, and age-decade group was calculated on final recall results. The analysis revealed a main effect of list, $F(2, 464) = 22.30$, $MSE = 249.33$, $p < .001$, $\eta_p^2 = .088$, a main effect of age-decade group, $F(3, 232) = 20.93$, $MSE = 1221.61$, $p < .001$, $\eta_p^2 = .213$, and an interaction between list and practice condition, $F(2, 464) = 24.03$, $MSE = 249.33$, $p < .001$, $\eta_p^2 = .094$. No other main or interaction effects were significant, all $Fs < 1$, $ps \geq .465$. Thus, no significant influence of age-decade group on testing effects was observed in the overall analysis. In the next step, the final recall results were analyzed separately for the three lists.

List 3. Final List-3 recall results by and large mirrored the immediate List-3 recall results (see Figure 3A). A two-way ANOVA of practice condition and age-decade group showed main effects of practice condition, $F(1, 232) = 9.13$, $MSE = 593.28$, $p = .003$, $\eta_p^2 = .038$, and age-decade group, $F(3, 232) = 17.16$, $MSE = 593.28$, $p < .001$, $\eta_p^2 = .182$, but no interaction between the two, $F(3, 232) < 1$. Participants recalled more items in the testing condition than in the restudy condition (44.7% vs. 35.2%).

The 40s group (56.5%) recalled more items than the 50s (43.5%), 60s (33.5%), and 70s (26.5%) groups, $ps_{\text{tukey}} \leq .002$; the 50s group recalled more items than the 70s group, $p_{\text{tukey}} < .001$. Bayesian analysis showed that the data were 15.31 times more likely under the two main-effects model than under the model that added the interaction. Consistent with the immediate List-3 recall results above, the final List-3 recall results indicate a benefit of testing over restudy that did not depend on age-decade group.

List 2. Final List-2 recall is depicted in Figure 3B. A two-way ANOVA of practice condition and age decade revealed a main effect of age-decade group, $F(3, 232) = 14.36$, $MSE = 629.09$, $p < .001$, $\eta_p^2 = .157$, but neither a main effect of practice condition, $F(1, 232) = 1.11$, $MSE = 629.09$, $p = .292$, $\eta_p^2 = .005$, nor an interaction between the two variables, $F(3, 232) < 1$. Post hoc tests showed that the 40s (44.0%) and 50s (37.0%) groups recalled significantly more items than the 60s (23.2%) and 70s (17.3%) groups, $ps_{\text{tukey}} \leq .014$. Consistently, Bayesian analysis revealed that the data were best described by the model with the single main effect of age-decade group. Compared with this model, the data were 4.19 times less likely under the two main-effects model and 77.30 times less likely under the two-effects-plus-interaction model.

List 1. Final List-1 recall results are shown in Figure 3C. A two-way ANOVA with practice condition and age-decade group showed main effects of practice condition, $F(1, 232) = 12.46$, $MSE = 497.90$, $p < .001$, $\eta_p^2 = .051$, and age-decade group, $F(3, 232) = 13.72$, $MSE = 497.90$, $p < .001$, $\eta_p^2 = .151$, but no interaction between the two variables, $F(3, 232) < 1$. Indeed, participants recalled more items in the restudy condition than in the testing condition (40.1% vs. 29.9%). The 40s (46.2%) and 50s (41.7%) groups recalled significantly more items than the 60s (28.7%) and 70s (23.5%)

³ Regarding prior-list intrusions in the immediate List-3 recall test, a three-way ANCOVA with practice condition (testing vs. restudy) and age-decade group (40s vs. 50s vs. 60s vs. 70s) and the three covariates (years of formal education, performance on the backward-counting task, performance on the MWT-B) revealed a main effect of practice condition, $F(1, 229) = 9.37$, $MSE = 1.11$, $p = .002$, $\eta_p^2 = .039$, but no significant main effect of age-decade group, $F(3, 229) = 1.95$, $MSE = 1.11$, $p = .122$, $\eta_p^2 = .025$, and no significant interaction, $F(3, 229) < 1$. Participants' years of formal education had a negative influence on the number of prior-list intrusions, $F(1, 229) = 5.76$, $MSE = 1.11$, $p = .017$, $\eta_p^2 = .025$, whereas participant performance on the backward-counting task and the MWT-B did not influence the number of prior-list intrusions, $Fs(1, 229) \leq 1$. The correlation between participants' years of formal education and the number of prior-list intrusions, with age partialled out, was $r(237) = -.114$, $p = .079$.

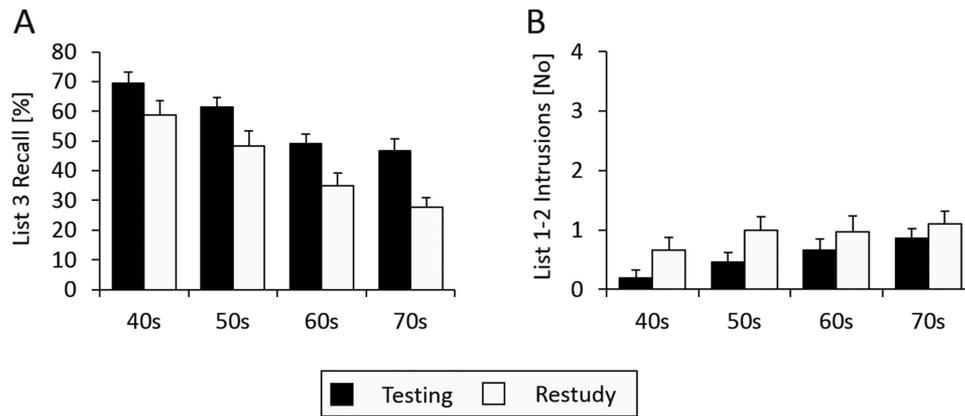


Figure 2. Results of the immediate List-3 recall test. (A) Correct recall rates as a function of practice condition (testing, restudy) and age-decade group (40s, 50s, 60s, 70s). (B) Number of prior-list intrusions as a function of condition and age-decade group. Error bars represent *SEMs*.

groups, $p_{\text{Tukey}} \leq .009$. Bayesian analysis showed that the data were 12.40 times more likely under the two main-effects model than under the model that added the interaction, thus indicating strong evidence against the interaction effect. These results suggest a reliable recall benefit of restudy over testing for the List-1 item material, unaffected by age-decade group.

Discussion

This study examined the late developmental trajectory of the forward effect of testing from middle adulthood to older age. The results showed that the forward effect is about equally present in four age-decade groups of participants ranging from 40 to 79 years of age: Immediate testing of preceding Lists 1 and 2 enhanced correct recall of the critical List 3 and reduced the number of prior-list intrusions in the immediate List-3 recall test, and did so similarly in each decade group. Analogous recall results were observed in the final List-3 recall test with respect to correct recall. The results suggest that the forward effect of testing is not only present in young adults, but that testing also enhances memory and learning of subsequent material in middle-aged and older adults. Following Szpunar et al. (2008), these results

suggest that testing of previously studied (nontarget) material can reduce PI for subsequently studied (target) material, and that this PI reduction is similarly efficient in middle-aged and older adults.

The finding that older adults show efficient PI reduction is consistent with findings from LMDF research. In the LMDF task, in which participants are cued to forget a first list and remember a second list instead, the forget cue impairs memory for the first list and enhances memory for the second list. Developmental studies have shown that the enhancement effect in LMDF is about equally present in young adults (approx. 20 years of age), young-older adults (approx. 70 years of age), and old-older adults (approx. 85 years of age; Aslan & Bäuml, 2013; Zellner & Bäuml, 2006). Following the view that the enhancement effect reflects PI reduction (Pastötter et al., 2017; Sahakyan et al., 2013), this work indicates that older adults are able to efficiently reduce PI in the LMDF task.

The present findings suggest a parallel between the forward effect of testing and the enhancement effect in LMDF, indicating that older adults' ability to reduce PI is not restricted to the LMDF task, but generalizes to the present interim-testing task. Future work may

Table 2
Results in the Immediate Recall Tests as a Function of Age-Decade Group and Practice Condition

Independent variables		Immediate recall rates			Intrusions
Age-decade group	Condition	List 1	List 2	List 3	List 3
40s	Testing	77.00 (2.96)	64.67 (4.03)	69.33 (3.86) [†]	.20 (.07) [*]
	Restudy			58.67 (4.84)	.67 (.21)
50s	Testing	68.00 (3.36)	58.33 (4.21)	61.33 (3.28) [*]	.47 (.15) [†]
	Restudy			48.33 (4.98)	1.00 (.25)
60s	Testing	63.67 (3.13)	51.00 (2.77)	49.00 (3.40) [*]	.67 (.18)
	Restudy			35.00 (4.20)	.97 (.27)
70s	Testing	55.00 (3.31)	48.00 (4.56)	46.67 (4.13) ^{**}	.87 (.16)
	Restudy			27.67 (3.13)	1.10 (.22)

Note. Pairwise *t* tests (two sided, uncorrected); reliable differences between conditions (testing vs. restudy); means and standard errors (in brackets).

[†] $p < .10$. ^{*} $p < .05$. ^{**} $p < .01$.

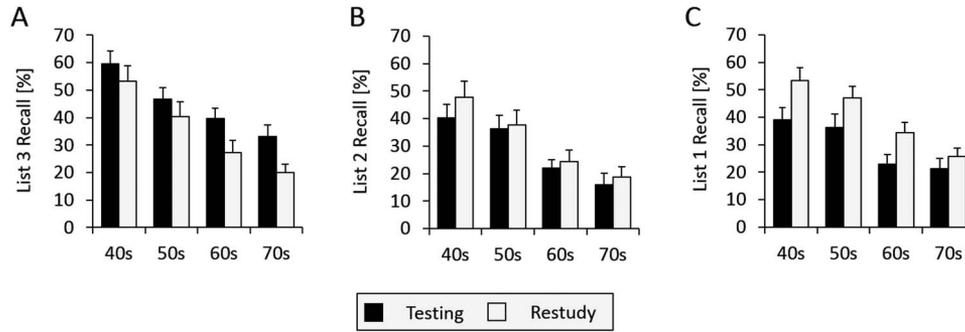


Figure 3. Results of the three final recall tests. (A) Correct List-3 recall rates as a function of practice condition (testing, restudy) and age-decade group (40s, 50s, 60s, 70s). (B) Correct List-2 recall rates as a function of practice condition and age-decade group. (C) Correct List-1 recall rates as a function of practice condition and age-decade group. Error bars represent *SEMs*.

examine to what extent the forward-testing effect and the LMDF-enhancement effect are functionally and neurally equivalent. Elsewhere, it has been argued that the forward effect may be mediated by some form of self-induced forget instruction, inducing participants to think that the practiced information is no longer needed and thus can be forgotten (Szpunar, McDermott, & Roediger, 2007). Others have argued that the forward effect of testing and LMDF are mediated by the same context-change mechanism, with both testing and a forget cue inducing a change in individuals' internal context (Divis & Benjamin, 2014). Whether the suggested equivalences really hold is far from clear, however (see Abel & Bäuml, 2016, 2017), and further research is required to clarify the relation between the forward effect and LMDF.

In the literature, both encoding and retrieval factors have been suggested to contribute to PI reduction and the forward effect of testing in younger adults. Specifically, testing has been argued to improve temporal discrimination between preceding and critical lists at test (Bäuml & Kliegl, 2013; Szpunar et al., 2008) and to enhance attentional encoding of the critical list items (Pastötter et al., 2011). Both improved encoding and improved retrieval may also have contributed to the observed forward effect in older adults, although this proposal awaits more direct examination in the future. For instance, on the basis of the neurocognitive study by Pastötter et al. (2011), which showed that alpha brain oscillations are a neural marker of improved encoding after testing in younger adults, future work may examine the effects of testing on oscillatory brain activity in older adults. Such work may also address older adults' retrieval abilities in the present task. Following Bäuml and Kliegl (2013), response-latency analysis of recall data may be employed to examine whether, as in young adults, retrieval practice enhances older adults' temporal differentiation between preceding and critical lists at test, thus creating more focused search of the critical items. Together, such work may also shed new light on the inhibitory deficit account, which assumes that aging is associated with decreased ability to ignore nontarget information and focus attention on the target information (Hasher & Zacks, 1988). Although the present findings are consistent with the inhibitory deficit account of aging, they also suggest that retrieval practice can increase older adults' ability to ignore nontarget information and focus attention on target information. Future work relating electroencephalographic alpha oscillations and

response latencies to the forward effect of testing in older adults may test the contribution of encoding and retrieval factors to inhibitory deficits in aging in more detail.

The present results, together with previous work, suggest that the forward effect of testing is a robust phenomenon present in both younger and older adults. With regard to younger adults, previous studies have found that interim testing eliminates PI buildup across item lists (Aslan & Bäuml, 2016; Pastötter et al., 2011; Szpunar et al., 2007, 2008), that is, preceding and critical lists have shown the same immediate recall rates in the testing condition. In contrast, in the present study, immediate List-1 recall was reliably higher than immediate List-2 and List-3 recall. This finding suggests that testing may have reduced, but not eliminated PI buildup in middle-aged and older adults. Together with findings from the previous work, this finding may reflect a dissociation between younger and older adults, but further work is required to examine whether PI buildup is really different between age groups.⁴ In addition, with regard to older adults, future researchers may examine whether the forward effect generalizes from young-older adults (approx. 70 years of age) to old-older adults (approx. 85 years of age). On the basis of the finding that old-olders show efficient PI reduction in the LMDF task (Aslan & Bäuml, 2013), one may expect such generalization.

With regard to final recall testing, the results show that participants recalled more of the restudied than of the retrieval-practiced List-1 items and the same number of restudied and retrieval-practiced List-2 items, independent of age-decade group. Regarding List 1, the finding that restudied items are better recalled than retrieval-practiced items is consistent with the testing-effect literature, showing that the backward-testing effect is most prominent when the final criterion test is administered after a relatively long

⁴ A study by Pastötter, Engel, & Frings (2018) recently examined the forward effect of testing on immediate List-3 recall in 240 younger university students when employing an experimental setup that was similar to the present one. Effect size calculation of Cohen's *d* for nonrepeated measures revealed a medium effect size for the forward effect regarding both correct recall, $d = .614$, and number of prior-list intrusions, $d = .396$. In the present study, similar effect sizes for correct recall, $d = .583$, and number of prior-list intrusions, $d = .359$, were observed for middle-aged and older adults from the community, suggesting that the forward effect in target-list recall may be similar in size in younger and older adults.

retention interval (e.g., days or weeks), but is often eliminated or even reversed when the final test is administered after a relatively short retention interval (e.g., several min; see Hogan & Kintsch, 1971; Kornell, Bjork, & Garcia, 2011; Roediger & Karpicke, 2006). The present finding of no interaction between age and practice condition in List-1 recall is also consistent with a recent study by Coane (2013), which examined the effects of retrieval practice (with feedback) versus restudy after short- and long-retention intervals in younger adults (approx. 20 years of age) and older adults (approx. 65 years of age). She found a reliable backward effect on the long-retention interval (2 days) but no effect of practice condition on the short-retention interval (10 min). In particular, she found that age did not interact with practice condition, suggesting that older adults benefited (in the long term) as much as younger adults from retrieval practice (see also Tse, Balota, & Roediger, 2010, for a possible moderation of feedback backward-testing effect in older adults). Regarding List 2, neither an effect of practice condition nor an interaction between age and practice condition was observed. Because final List-2 recall provides a mixed measure of the backward and forward effects, these findings are in line with the final List-1 and List-3 recall results, which provide rough measures of the backward and forward effects, respectively (see Method section).

As mentioned above, delay plays a critical role for the backward-testing effect, often reversing a benefit for restudy in the short term into a benefit for retrieval in the long term (e.g., Hogan & Kintsch, 1971; Kornell et al., 2011; Roediger & Karpicke, 2006). In contrast, delay does not seem to play such a role for the forward-testing effect. As recently reported by Chan, Manley, Davis, and Szpunar (2018), the forward effect can still be present when the List-3 recall test is delayed by 25 min, at least in young adults. Future work may address the permanence of the effect and examine whether it varies with subjects' age levels. Such experiments would offer an even stronger test of the role of age in forward-testing effects.

To conclude, as documented in numerous studies in the literature, older adults show memory impairment over a wide range of episodic memory tasks. Retrieval-practice effects, as they have been studied in the testing-effects literature, seem to be an exception to this "rule." Prior work has shown that retrieval practice on studied items improves later recall of the practiced items in older adults similar to how it does in younger adults. The present study shows that, in both middle-aged and older adults, retrieval practice can improve recall of subsequently studied material. These findings on the backward and forward testing effects indicate that retrieval practice can be a very powerful tool to delimit memory impairment in older age.

References

- Abel, M., & Bäuml, K.-H. T. (2016). Retrieval practice can eliminate list method directed forgetting. *Memory & Cognition*, *44*, 15–23. <http://dx.doi.org/10.3758/s13421-015-0539-x>
- Abel, M., & Bäuml, K.-H. T. (2017). Testing the context-change account of list-method directed forgetting: The role of retention interval. *Journal of Memory and Language*, *92*, 170–182. <http://dx.doi.org/10.1016/j.jml.2016.06.009>
- Aslan, A., & Bäuml, K.-H. T. (2013). Listwise directed forgetting is present in young-old adults, but is absent in old-old adults. *Psychology and Aging*, *28*, 213–218. <http://dx.doi.org/10.1037/a0031295>
- Aslan, A., & Bäuml, K. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science*, *19*, 992–998. <http://dx.doi.org/10.1111/desc.12340>
- Balota, D. A., Duchek, J. M., Sergent-Marshall, S. D., & Roediger, H. L., III. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging*, *21*, 19–31. <http://dx.doi.org/10.1037/0882-7974.21.1.19>
- Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, *68*, 39–53. <http://dx.doi.org/10.1016/j.jml.2012.07.006>
- Biss, R. K., Campbell, K. L., & Hasher, L. (2013). Interference from previous distraction disrupts older adults' memory. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, *68*, 558–561. <http://dx.doi.org/10.1093/geronb/gbs074>
- 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>
- Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language*, *102*, 83–96. <http://dx.doi.org/10.1016/j.jml.2018.05.007>
- Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory & Cognition*, *2*, 95–100. <http://dx.doi.org/10.1016/j.jarmac.2013.04.001>
- Crowder, R. G. (1976). *Principles of learning and memory*. New York, NY: Wiley.
- Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*, 1049–1062. <http://dx.doi.org/10.3758/s13421-014-0425-y>
- Duyck, W., Desmet, T., Verbeke, L. P., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments & Computers*, *36*, 488–499. <http://dx.doi.org/10.3758/BF03195595>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. <http://dx.doi.org/10.3758/BF03193146>
- Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology*, *60*, 991–1004. <http://dx.doi.org/10.1080/17470210600823595>
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193–225. [http://dx.doi.org/10.1016/S0079-7421\(08\)60041-9](http://dx.doi.org/10.1016/S0079-7421(08)60041-9)
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, *10*, 562–567. [http://dx.doi.org/10.1016/S0022-5371\(71\)80029-4](http://dx.doi.org/10.1016/S0022-5371(71)80029-4)
- Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subjective reports of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 686–700. <http://dx.doi.org/10.1037/0278-7393.27.3.686>
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory* (pp. 487–514). Oxford, UK: Academic Press. <http://dx.doi.org/10.1016/B978-0-12-809324-5.21055-9>
- 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). San Diego, CA: Academic Press.

- Kliegl, O., Pastötter, B., & Bäuml, K.-H. T. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1778–1789. <http://dx.doi.org/10.1037/xlm0000096>
- 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>
- Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B*. Balingen, Germany: Spitta-Verlag.
- Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology*, *50*, 994–1000. <http://dx.doi.org/10.1037/a0035202>
- Meyer, A. N., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*, 142–147. <http://dx.doi.org/10.1037/a0030890>
- Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology: Series B. Psychological Sciences and Social Sciences*, *72*, 82–90. <http://dx.doi.org/10.1093/geronb/gbw066>
- Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, Article 286. <http://dx.doi.org/10.3389/fpsyg.2014.00286>
- Pastötter, B., Engel, M., & Frings, C. (2018). The forward effect of testing: Behavioral evidence for the reset-of-encoding hypothesis using serial position analysis. *Frontiers in Psychology*, *9*, Article 1197. <http://dx.doi.org/10.3389/fpsyg.2018.01197>
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287–297. <http://dx.doi.org/10.1037/a0021801>
- Pastötter, B., Tempel, T., & Bäuml, K.-H. T. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology*, *8*, Article 2076. <http://dx.doi.org/10.3389/fpsyg.2017.02076>
- Pastötter, B., Weber, J., & Bäuml, K.-H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, *27*, 280–285. <http://dx.doi.org/10.1037/a0031797>
- Pauls, F., Macha, T., & Petermann, F. (2013). U-shaped development: An old but unsolved problem. *Frontiers in Psychology*. Advance online publication. <http://dx.doi.org/10.3389/fpsyg.2013.00301>
- 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>
- Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 59, pp. 131–189). New York, NY: Elsevier.
- Sumowski, J. F., Chiaravalloti, N., & Deluca, J. (2010). Retrieval practice improves memory in multiple sclerosis: Clinical application of the testing effect. *Neuropsychology*, *24*, 267–272. <http://dx.doi.org/10.1037/a0017533>
- Sumowski, J. F., Wood, H. G., Chiaravalloti, N., Wylie, G. R., Lengenfelder, J., & DeLuca, J. (2010). Retrieval practice: A simple strategy for improving memory after traumatic brain injury. *Journal of the International Neuropsychological Society*, *16*, 1147–1150. <http://dx.doi.org/10.1017/S1355617710001128>
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition*, *35*, 1007–1013. <http://dx.doi.org/10.3758/BF03193473>
- 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>
- Tse, C.-S., Balota, D. A., & Roediger, H. L., III. (2010). The benefits and costs of repeated testing on the learning of face–name pairs in healthy older adults. *Psychology and Aging*, *25*, 833–845. <http://dx.doi.org/10.1037/a0019933>
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*, 49–60. <http://dx.doi.org/10.1037/h0044616>
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538. <http://dx.doi.org/10.3758/BRM.41.2.534>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology: Part II. Example applications with JASP. *Psychonomic Bulletin & Review*, *25*, 58–76. <http://dx.doi.org/10.3758/s13423-017-1323-7>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: NCS Pearson.
- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1024–1039. <http://dx.doi.org/10.1037/0278-7393.19.5.1024>
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *npj Science of Learning*. Advance online publication. <http://dx.doi.org/10.1038/s41539-018-0024-y>
- Zellner, M., & Bäuml, K.-H. (2006). Inhibitory deficits in older adults: List-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 290–300. <http://dx.doi.org/10.1037/0278-7393.32.3.290>

Received April 26, 2018

Revision received August 28, 2018

Accepted September 6, 2018 ■