Journal of Applied Research in Memory and Cognition xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Retrieval practice and elaborative encoding benefit memory in younger and older adults

Jennifer H. Coane*

Colby College, United States

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form 3 April 2013
Accepted 11 April 2013
Available online xxx

Keywords:
Retrieval practice
Feedback
Aging
Deep processing
Memory

ABSTRACT

Retrieval practice has been identified as a powerful tool for promoting retention. Few studies have examined whether retrieval practice enhances performance in older adults as it does in younger adults. Younger and older adults learned unrelated word pairs and were administered a test after a short (10 min) and long (2 day) delay. Encoding condition was manipulated between subjects, with participants studying the pairs twice, studying them once and taking an immediate test with feedback, or encoding them twice under different deep encoding conditions. In both age groups, equivalent benefits of testing relative to restudy were found. Deep processing also improved memory relative to restudy, suggesting that one factor that might benefit retention is varying the type of encoding task (either by testing or by providing a different instructional manipulation) to increase the accessibility of cues. Retrieval practice can support older adults' memory and is a viable target for training.

Numerous studies demonstrate that retrieval practice benefits long-term retention relative to repeated study across a variety of materials (e.g., paired associates, prose passages, maps) and populations (e.g., middle school children, high school and college students; see Roediger & Karpicke, 2006a; Roediger, Agarwal, Kang, & Marsh, 2010; and Roediger, Putnam, & Smith, 2011, for reviews). In typical studies, participants learn some material, such as paired associates, and then are given another opportunity to study the material (repeated study condition) or take an immediate test (e.g., a cued recall test in which one item is presented and participants must retrieve the other item in the pair). In this study-test condition, accuracy or corrective feedback may or may not be provided. After a delay, a final test is administered to both groups. The consistent finding is better performance or reduced forgetting in the study-test condition relative to the repeated study condition after a delay.

Although the practical benefits of prior testing are well demonstrated, few studies have examined whether testing benefits a population with known memory deficits such as the elderly. Clearly, demonstrating a testing effect in older adults would have broad implications for training and rehabilitation purposes as well as understanding the factors involved in the testing effect.

1. Memory deficits in aging

Older adults typically perform worse than younger adults on tests that tap episodic memory (Balota, Dolan, & Duchek, 2000). Aging effects are especially pronounced in tasks that require associating units of information, such as face-name pairs (e.g., Logan & Balota, 2008) or word pairs (e.g., Naveh-Benjamin, 2000), or remembering the source of information (e.g., Hashtroudi, Johnson, & Chrosniak, 1989).

According to the *elaboration deficit* hypothesis (Kausler, 1982), older adults tend not to engage in self-initiated elaboration strategies during encoding. Furthermore, they may not use such strategies effectively even after training (Nyberg, 2005). Similarly, Craik (Craik, 1986; Craik & Rabinowitz, 1985) argued that older adults are less likely to spontaneously engage in effortful processing and online monitoring of learning. In support of these accounts, when tests provide more environmental support or richer cues, older adults show reduced deficits. Within the levels-of-processing framework (Craik & Lockhart, 1972) greater degrees of meaningful elaboration or processing typically result in better retention on recall and recognition tests. Because older adults do not elaborate spontaneously, they typically benefit from directed processing instructions (e.g., Erber, Galt, & Botwinick, 1985; Erber, Herman, & Botwinick, 1980; Rabinowitz, Craik, & Ackerman, 1982). The *associative deficit* hypothesis (Naveh-Benjamin, 2000) attributes age-related memory declines to selective difficulties in binding items either to one another, as in a paired associate task, or to the source or learning episode. Additional explanations attribute

---

* Correspondence address: Department of Psychology, Colby College, Waterville, ME 04901, United States. Tel.: +1 207 859 5556.
E-mail address: jhcoane@colby.edu

2211-3681/$ – see front matter © 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jarmac.2013.04.001

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001

age-related declines to deficits in inhibitory processes (Hasher & Zacks, 1979), or less effective suppression of irrelevant information.

Because of age-related memory declines, the effectiveness of cognitive training to improve objective memory performance and reduce subjective memory complaints has been researched extensively (e.g., Hertzog, Kramer, Wilson, & Lindenburger, 2009; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009; Rebok, Carlson, & Langbaum, 2007). Cognitive training can include training in one or more mnemonics (e.g., method of loci, face-name pairs) or a variety of strategies and techniques. Overall, training appears to be effective for the specific skill or strategy (e.g., Floyd & Scogin, 1997; Verhaeghen, Marcoen, & Goossens, 1992), although there is limited evidence for transfer to everyday living activities (e.g., Nyberg, 2005; West, Welch, & Yassuda, 2000) and some strategies might not generalize to all situations. Because retrieval practice consistently yields mnemonic benefits in young adults, it is important to demonstrate the effectiveness of testing in older populations. Furthermore, because testing is effective with a variety of materials, from word lists and prose passages (e.g., Roediger & Karpicke, 2006a), to visually complex stimuli (e.g., Kang, 2010), this strategy has the potential to support older adults' memory in numerous ways.

1.1. Testing effects in aging

Relatively few studies have directly examined whether testing benefits older adults' memory performance. "Interim" tests, i.e., tests interleaved with study events, can improve performance (e.g., Kausler & Wiley, 1991; Rabinowitz & Craik, 1986). Such improvements might be due to increasing motivation or metacognition, encouraging intentional encoding, or providing feedback (Rogers & Gilbert, 1997). These studies, however, generally had short retention intervals of only a few minutes.

Recently, Meyer and Logan (2013), using prose passages, found that college-aged individuals (students and non-students) and older adults (aged 55–65) showed robust and equivalent benefits of prior testing after short (5 min) and long (2 day) delays. Participants were given overall accuracy feedback (i.e., how many items were answered correctly on the initial test). The same items were presented in both tests, as multiple choice questions initially, and as cued recall items on the final test. As the authors acknowledge, this might have increased attention to the subset of items on the test. Tse, Balota, and Roediger (2010) also found a testing effect in old adults using face–name pairs, but only when feedback was provided during the initial test. Without feedback, young–old adults (mean age 72 years) recalled the same amount of information in the repeated study and in the repeated test conditions, whereas old–old adults (mean age 81 years) actually showed a benefit of repeated study. However, the retention interval in Tse et al. was relatively short, at 1.5 h, and participants only studied eight face–name pairs. Thus, further research with different materials and retention intervals is needed in this area. Because older adults tend to show more marked deficits in associative learning tasks, it is important to provide additional evidence of potential benefits of testing. Tse et al. did provide evidence in their face–name task, but the retention interval was substantially shorter than the one used in other studies. Thus, in the present study, associative learning, assessed through paired associates, was examined following a short (10 min) and long (2 day) retention interval.

1.2. Theoretical accounts of the testing effect

According to the *processing match* account, there is greater overlap in the type of processing between successive tests than in repeated study and a later test. Thus, as the similarity between tests increases, so should performance. However, several studies have suggested that initial recall (i.e., short answer) tests result in better retention even when the final test is a recognition (i.e., multiple choice) test (e.g., Kang, McDermott, & Roediger, 2007; see also Balota & Neely, 1980). Thus, similarity of operations or processing match cannot totally account for the testing benefit.

According to the *semantic mediator* hypothesis (Pyc & Rawson, 2010), compared to restudy, testing results in qualitatively different (i.e., richer, more elaborative) mediating information that links cues and targets. In support of this hypothesis, Carpenter (2011) found that participants recalled more targets (e.g., *child*) when the test cue (e.g., *father*) was semantically related to the study cue (e.g., *mother*) than when an independent cue (e.g., *birth*) was given at test. Carpenter suggested that testing enhances the semantic content of information in long-term memory. When targets are weakly related to their cues, more elaborative processing occurs, thus increasing the availability of semantic mediators. Similarly, according to the *elaborative retrieval* hypothesis proposed by Carpenter (2009), retrieval efforts are more likely to result in elaboration as individuals search through memory. Such a search is likely to activate related items in semantic memory, thus creating a richer memory trace (e.g., Collins & Loftus, 1975).

1.3. The current study

In the present study, in addition to a repeated study condition and a study-test condition, a *Deep Processing* condition was included in which participants encoded the items incidentally in two semantically rich manners. The two encoding conditions required finding similarities between the words in the pair and generating a mental image connecting the items (Pecher & Raaijmakers, 2004). Deep processing influences later retention (e.g., Craik & Tulving, 1975) and was expected to improve performance relative to repeated study. Because older adults are less likely to spontaneously engage in elaborative encoding, older adults in particular might benefit from an instructional manipulation encouraging deep processing and elaboration. The tasks varied across presentations to avoid participants simply retrieving prior operations (e.g., generating the same mental image) at the second presentation. If elaboration of richer semantic traces contributes to the testing effect, performance in this condition should be similar to the study-test condition. However, if the testing effect is due to retrieval-specific factors, this condition should result in worse performance than the retrieval practice condition.

As suggested by Carpenter (2009, 2011) and Pyc and Rawson (2010), repeated testing promotes elaboration of a richer semantic context. Because older adults frequently show age-related deficits in spontaneous elaborative processing (e.g., Craik, 1986), one might expect that retrieval practice, by increasing elaborative or semantic processing would help attenuate some of the age-related declines in retention. However, if the type of elaborative processing involved in retrieval practice depends on spontaneous, rich encoding, it is possible older adults would not benefit from testing. Furthermore, because of associative deficits in aging (e.g., Naveh-Benjamin, 2000), testing may fail due to insufficient learning in the initial study phase. Thus, retrieval practice might be less effective in an older population because of failures in elaboration during the retrieval process or because of associative deficits during the encoding phase.

2. Method

2.1. Participants

Sixty-nine young adults from Colby College's research participant pool and from the Waterville Community and 70 older adults

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001

Table 1
Demographic characteristics of young ($N = 29$) and old ($N = 60$) adults (standard error of the mean in parentheses).

	Young	Old	*p*-Value
Age (years)	19.78 (0.17)	67.44 (0.56)	<0.001
Education (years)	13.64 (0.16)	15.79 (0.36)	<0.001
O-Span (# correct)	28.38 (0.94)	21.17 (1.01)	<0.001
Shipley (# correct)	32.11 (0.69)	34.92 (0.52)	0.002
DSST (s)	65.97 (2.24)	52.44 (1.10)	<0.001
Trails B (s)	48.79 (3.05)	81.76 (5.72)	<0.001

from the Waterville community participated. Data from five older adults and one younger adult were excluded because of computer failures or failure to return for the second session. Data from two additional older adults were omitted for failure to follow instructions. See Table 1 for demographic characteristics of the remaining participants. Participants were administered a battery of cognitive tasks assessing working memory (Operation Span Task; Unsworth, Heitz, Schrock, & Engle, 2005), vocabulary (Shipley, 1940), executive control (Trails B; Reitan, 1958), and processing speed (Digit Symbol Substitution Test, DSST; Wechsler, 1997). Due to experimenter error, cognitive battery scores are available for some of the young adult sample ($n = 29$) and 60 older adults. Twenty older adults were in the Study-Test condition, 21 in the Study–Study condition, and 22 in the Deep Processing condition. Twenty-four young adults were assigned to the Study-Test condition, 21 to the Study–Study condition, and 23 to the Deep Processing condition. Older adult participants were compensated with $15 and college students were given course credit or $15. Colby College's Review Board approved the study.

2.2. Materials

The materials consisted of 44 word pairs (e.g., *demon-dark*, *horse-jumped*) that were unrelated but could be connected easily through imagery or sentence generation.[1] The 44 pairs were divided into two sets of 22 for counterbalancing.

The automated version of the Operation Span Task (Unsworth et al., 2005) and a computerized version of the Shipley (1940) task were used. The Trails B and DSST tasks were administered using paper versions.

2.3. Procedure

The experiment consisted of four phases: An initial learning phase (Block 1), a second learning phase (Block 2), a test after 10 min, and a final test after 2 days. In the Study–Study condition, participants were instructed to study the pairs any way they chose. Each pair was presented on the computer monitor in white on a black background for 5 s, with a 500 ms inter-stimulus interval (ISI). After a first presentation of all 44 pairs, the pairs were presented again in Block 2 in a new random order, with the same instructions. In the Deep Processing condition, in Block 1, participants were told to try to find similarities between the pairs and to "try to find something these two items have in common." They were told there were no right or wrong answers. In Block 2, participants were instructed to "create a mental image that combines the two words." No overt response was required in either task. In both cases, the pairs were presented for 5 s each, with a 500 ms ISI. In the Study-Test condition, Block 1 was identical to Block 1 in the Study–Study condition. In Block 2, the first item in each pair (the cue) was presented with three question marks (e.g., *demon* -???, *horse* -???) and participants were asked to type the word it had been paired with in the study phase. Participants were encouraged to try to retrieve the correct item, or to enter "XXX" if they could not remember the item. Participants were given unlimited time to respond, to reduce performance or testing anxiety in older adults (e.g., Earles, Kersten, Mas, & Miccio, 2004; Henkel, 2007; Hess & Hinson, 2006). One concern was that older participants would not engage in retrieval if they felt they did not have enough time. Differences in typing speed were also a concern (several older adults reported limited experience with computers). After responding, the correct answer was displayed on the screen until participants pressed a key to advance to the next trial (i.e., feedback presentation was self-paced).

Following Block 2, participants completed an unrelated filler task (i.e., Sudoku puzzles). After 10 min, the instructions for the immediate test were presented. The instructions were the same as in the Test phase in the Study-Test condition. Twenty-two of the 44 pairs were presented in random order. There was no time limit and no feedback was provided. After completing the test, participants were reminded to return 2 days later. No specific instructions were given at that time.

Two days later, upon their return to the lab, participants were given the same test instructions and the remaining 22 cue words. Upon completion, they were debriefed and compensated. The cognitive battery was administered either before the Day 1 encoding or after the test on Day 2 to accommodate participants' schedules and to minimize fatigue on the first day.

3. Results

Responses were scored as correct if the target generated was identical to the studied target or if minor morphological variations occurred (e.g., *kill* instead of *killed* in response to *turkey*). Unless otherwise noted, all significant effects are $p < 0.01$.

Correct responses were submitted to a 2 (age) × 3 (condition: Study–Study, Study-Test, Deep Processing) × 2 (test delay: 10 min vs. 2 days) mixed ANOVA (see Fig. 1). Age and condition were between-subjects factors and test delay was a within-subjects factor. The effect of age was reliable, $F(1,125) = 13.46$, $\eta_p^2 = 0.10$. Older adults recalled fewer targets ($M = .40$, $SEM = 0.022$) than younger adults ($M = 0.51$, $SEM = 0.021$). Importantly, however, age did not

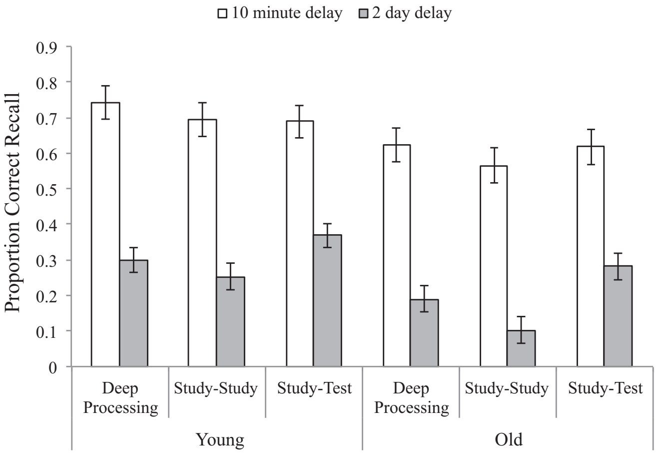

Fig. 1. Mean correct recall on the tests after 10 min and 2 days as a function of age and condition. Error bars represent the standard error of the mean.

---

[1] One-hundred and forty-seven word pairs (e.g., *demon-dark*, *horse-jumped*; Maddox, Balota, Coane, & Duchek, 2011) were initially selected. In pilot testing, 14 undergraduate participants rated the pairs on the ease of finding similarities between the words ($n = 7$) or forming a mental image connecting the words ($n = 7$). Both rating tasks were on a scale of 1 (very hard/impossible) to 5 (very easy). The 44 pairs with the highest ratings on both scales were selected. The average ratings were 4.45 (SEM = 0.034) for the ease of forming a mental picture and 4.25 (SEM = 0.054) for the ease of finding similarities.

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001

interact with any other factors, all *F*s < 1. The effect of delay was reliable, $F(1,125) = 575.42$, $\eta_p^2 = 0.82$. Correct recall declined from the first test ($M = 0.66$, $SEM = 0.02$) to the second test after 2 days ($M = 0.25$, $SEM = 0.015$). The main effect of condition approached conventional significance levels, $F(2,125) = 2.76$, $p = 0.067$, $\eta_p^2 = 0.4$.

Importantly, the interaction between delay and condition was significant, $F(2,125) = 5.36$, $\eta_p^2 = 0.08$. On the test administered after 10 min, there were no significant differences between conditions, $F < 1$. However, after a 2-day delay, an effect of condition emerged, $F(2,128) = 8.02$, $\eta_p^2 = 0.11$. Participants who took a test immediately after encoding outperformed those who did not. Specifically, a benefit of testing clearly emerged relative to repeated study, $t(84) = 3.93$, and relative to deep processing, $t(87) = 2.09$, $p = 0.04$. The effect was reliable in both age groups, relative to repeated study and deep processing, all $ps < 0.05$. Moreover, the Deep Processing condition yielded higher recall on the delayed test than the Study–Study condition, $t(85) = 1.99$, $p = 0.049$. Thus, compared to repeated studying, taking a test yields a large benefit on long-term retention, and deep processing yields a smaller, but reliable, benefit as well.

The Study-Test group might have benefitted from the fact that the initial retrieval event in Block 2 was self-paced, whereas the presentation rate in Block 2 in the other two groups was fixed at 5 s. Median RTs in the Study-Test group were 7739 ms for older adults and 4274 for younger adults. When median RTs were entered as a covariate in the analyses with age, delay, and condition, the interaction between delay and condition was still significant, $F(2,123) = 3.23$, $p = 0.04$, $\eta_p^2 = 0.05$, suggesting that exposure time was not uniquely responsible for the benefits of retrieval practice.

To assess performance on the test administered during the encoding phase in the Study-Test condition and the extent to which participants benefitted from feedback, an ANOVA with age and time of test as factors (i.e., the test occurring during Block 2 of the encoding phase and after the 10 min delay) was conducted. Only the effect of test was significant, $F(1,43) = 99.54$, $\eta_p^2 = 0.70$. Performance increased from an average of 0.46 ($SEM = 0.03$) on the test administered during Block 2 to 0.65 ($SEM = 0.034$) on the 10-min test. Neither the effect of age nor the age by test interaction was significant, both *F*s < 2.2, *p*s > 0.15. Thus, younger and older adults benefited equally from the initial test and the feedback, as demonstrated by the substantial boost in performance from the initial test to the test administered 10 min later.

There is an apparent discrepancy in the data between the non-significant effect of age in the Study-Test group reported above and the overall effect of age in the omnibus ANOVA. Even when the 2-day test was included in the analysis of the Study-Test condition, the effect of age failed to emerge ($p = 0.20$, $\eta_p^2 = 0.04$), although numerically younger adults did outperform older adults. It should be noted that the effect size of the age by condition interaction in the omnibus ANOVA was very small ($\eta_p^2 = 0.03$). Separate ANOVAs indicate that the effect of age was significant in both the Study–Study and Deep Processing groups, $ps < 0.035$ ($\eta_p^2 = 0.19$ and $\eta_p^2 = 0.10$, respectively). These analyses, in sum, suggest that age effects could be detected in the conditions in which they were present and that testing might moderate such age effects, although clearly further research is needed.

A follow-up analysis was conducted to test whether individual differences moderated the effects of retrieval practice on recall. This analysis included performance on the four cognitive tasks (O-Span, Trails B, Shipley, and DSST) as covariates, as well as two-way and three-way interactions between the cognitive tasks, time, and condition. None of the cognitive tasks' three-way interactions with time and condition were statistically significant (all $ps > 0.07$), indicating that cognitive performance did not moderate the effects of condition on recall.

4. Discussion

Consistent with prior studies (e.g., Meyer & Logan, 2013; Tse et al., 2010), initial testing enhanced retention relative to repeated study on the delayed test. Age did not interact with delay or condition, suggesting that older adults benefitted as much as younger adults from an initial encoding session that included a test with feedback. Meaningful and variable encoding also promoted better retention than repeated study on the delayed test. Importantly, a test with feedback resulted in the highest levels of performance on the final test, demonstrating the efficacy of retrieval practice on retention and confirming that processes specific to retrieval and to processing of corrective feedback are key to the benefits of testing, and that these processes produce more durable traces than richness of encoding generated by semantic or other processing.

Several accounts of age-related memory deficits attribute older adults' poor performance to impoverished encoding or difficulty using effective strategies (e.g., Craik, 1986; Nyberg, 2005). As proposed by Carpenter (2011) and Pyc and Rawson (2010), testing increases the degree of elaborative processing during the retrieval process. Whether such benefits would extend to a population that tends not to spontaneously engage in elaborative processing was examined here. Retrieval practice was indeed effective in an older population and, importantly, more effective than instructional manipulations aimed at promoting integrative semantic processing. These results suggest that retrieval practice itself promotes a form of elaborative processing that occurs during retrieval and may be more specific than the guided semantic processing occurring during encoding, which may be more general, especially in older adults (e.g., Rabinowitz & Craik, 1986). However, because the present study did not include a no-feedback condition, it cannot be determined to what extent the benefit of retrieval practice was influenced by additional processing associated with the corrective feedback. In fact, Tse et al. (2010) failed to observe a testing advantage in older adults in the absence of feedback. In the present study, both age groups did benefit substantially from feedback, suggesting that processing of corrective feedback, combined with retrieval practice, yields more durable learning than deep processing. Whether factors specific to the elaboration and search processes occurring during retrieval practice are sufficient for promoting long-term retention in the elderly or whether they depend on corrective feedback thus remains an open question.

Although the typical finding of better performance from repeated study on a test administered after 10 min was not observed (e.g., Roediger & Karpicke, 2006b), the benefit of prior testing clearly emerged after a 2-day delay. The inclusion of feedback might have contributed to the high performance in the Study-Test condition, by strengthening the memory representations of all tested items, including those that were not successfully retrieved originally (see Kornell, Bjork, & Garcia, 2011). Recently, Meyer and Logan (2013) reported a benefit of testing after as few as 5 min, suggesting the specific retention interval after which testing effects emerge is variable.

To examine whether older adults were less likely to engage in spontaneous elaboration (e.g., Craik, 1986), participants' responses to a questionnaire in which they described how they studied the pairs in the Study–Study condition were analyzed. Questionnaire data were available for 17 old and 18 young adults in the Study–Study condition. Strategy use was coded as deep or meaningful (e.g., forming a sentence or mental image) or shallow/no strategy (e.g., repetition). If participants reported using a deep encoding strategy, such as forming a sentence, their performance should have been enhanced relative to participants who did not use such a strategy. The likelihood of spontaneously engaging in meaningful processing was somewhat lower in older adults (53%, or 9 out of 17) than in younger adults (61%, or 11 out of 18), but

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001

this difference was not significant, $\chi^2 = 0.24$, $p = 0.62$. An ANOVA with age, strategy use, and test delay revealed only main effects of delay, $p < 0.001$, and of age, $p = 0.04$, consistent with the omnibus analyses. Importantly, strategy did not yield any effects nor did it interact with any other factors, all *F*s < 1, suggesting that meaningful processing, by itself, cannot explain the boost in performance of the deep processing group relative to the repeated study group. Thus, it seems that varying the strategies might have promoted the higher performance observed in the deep processing condition. Importantly, and as noted, the superior performance in the retrieval practice condition relative to the Deep Processing condition suggests that the elaborative processes occurring *during* retrieval attempts and in processing feedback are critical and that simply increasing the level of semantic processing at encoding is insufficient to match the power of testing as a mnemonic.

These results demonstrate the importance of retrieval practice in promoting long-term retention. Among the theoretical accounts of this phenomenon, the elaborative retrieval and semantic mediator accounts (Carpenter, 2011; Pyc & Rawson, 2010) attribute the testing benefit to increases in the richness of the traces as a result of retrieval attempts. Kornell et al. (2011), in their *distribution-based bifurcation model*, proposed that testing (especially without feedback) serves to selectively strengthen those items that can be retrieved in the initial test and that this selective strengthening is greater than the strengthening due to additional study. The current data are consistent with both of these approaches.

An alternative explanation that has not been extensively explored is that testing, relative to restudy, increases *encoding variability*. This account has been proposed to explain spacing effects (i.e., better performance on a test following spaced than massed practice; see Balota, Duchek, Sergent-Marshall, & Roediger, 2006). Briefly, this account, as applied to spacing, suggests that contextual elements are differentially available depending on the interval between successive presentations of an item. Successful retrieval depends on the overlap between contextual elements at encoding and at test. Variability increases with delay; thus, when there is greater variability in context, as in a delayed test, encoding conditions that encourage more variable and distinct encoding are less likely to be disrupted by the relatively new context at test. Compared to repeated study, testing is likely to engage different strategies and processes, and would thus be expected to result in greater variability of contextual and processing cues (e.g., Estes, 1955; see also McDaniel & Masson, 1985). In the present study, the Deep Processing condition included different processing instructions, which should promote more variability in contextual elements. The improved performance in the Deep Processing and Study-Test conditions thus might be consistent with an encoding variability account, although it is possible that performance in these two conditions is enhanced because of unrelated factors. Clearly, further studies to directly assess the potential contributions of encoding variability to the testing effect are needed.

5. Practical application

In closing, the applied benefits of this research merit highlighting. Older adults typically show impaired performance in basic memory tasks. Training interventions can improve performance. For example, a training program that included both an educational component (e.g., memory changes in age, types of memory) and a strategy use implementation (e.g., semantic elaboration, retrieval practice) improved subjective and objective memory performance in older adults relative to a no-contact control group (Troyer, 2001). This suggests that older adults would benefit both from testing and variable and meaningful encoding manipulations, although the present results indicate testing may be more beneficial. Other studies have found that training using effective mnemonics such as the method of loci improved performance in laboratory tasks but did not result in continued use following training (e.g., Scogin & Bienias, 1988). Some strategies, such as the method of loci, may not generalize to more complex materials or tasks. Thus, training programs that focus on strategies of broader applicability, such as retrieval practice or elaborative encoding, might be more beneficial. In fact, recent studies have suggested that testing also results in enhanced transfer to related or similar, but non-tested materials (e.g., Butler, 2010; Carpenter, 2012; Chan, 2010). Whether such advantages also extend to an aging population remains a question for future research.

Acknowledgements

Thanks to the student assistants, especially Anna Caron, Constance Jangro, Shannon Kooser, and Stephanie LaRose-Sienkiewicz. Thanks also to Dave Balota and Jan Duchek for helpful comments on earlier versions of the manuscript and to Chris Soto for statistical consulting.

References

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving, & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 395–410). Oxford: University Press.

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.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576–587.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*, 1118–1133.

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.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1547–1552.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*, 279–283.

Chan, J. C. K. (2010). Long-term effects of testing on the recall of non-tested material. *Memory*, *18*, 49–57.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix, & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409–422). Amsterdam, Holland: Elsevier.

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

Craik, F. I. M., & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology*, *40*, 309–315.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Earles, J. L., Kersten, A. W., Mas, B. B., & Miccio, D. M. (2004). Aging and memory for self- performed tasks: Effects of task difficulty and time pressure. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *59*, 285–293.

Erber, J. T., Galt, D., & Botwinick, J. (1985). Age differences in the effects of contextual framework and word-familiarity on episodic memory. *Experimental Aging Research*, *11*, 101–103.

Erber, J., Herman, T. G., & Botwinick, J. (1980). Age differences in memory as a function of depth of processing. *Experimental Aging Research*, *6*, 341–348.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145–154.

Floyd, M., & Scogin, F. (1997). Effects of memory training on the subjective memory functioning and mental health of older adults: A meta-analysis. *Psychology and Aging*, *12*, 150–161.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356–388.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, *4*, 106–112.

Henkel, L. A. (2007). The benefits and costs of repeated tests for young and older adult. *Psychology and Aging*, *22*, 580–595.

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2009). Enrichment effects on adult cognitive development. *Psychological Science in the Public Interest*, *9*, 1–65.

Hess, T. M., & Hinson, J. T. (2006). Age-related variation in the influences of aging stereotypes on memory in adulthood. *Psychology and Aging*, *21*(3), 621–625.

Kang, S. H. K. (2010). Enhancing visuo-spatial learning: The benefit of retrieval practice. *Memory and Cognition*, *38*, 1009–1017.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *The European Journal of Cognitive Psychology*, *19*, 528–558.

Kausler, D. H. (1982). *Experimental psychology and human aging*. New York: John Wiley and Sons.

Kausler, D. H., & Wiley, J. G. (1991). Effects of short-term retrieval on adult age differences in long-term recall of actions. *Psychology and Aging*, *4*, 661–665.

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.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal spaced retrieval practice in healthy young and older adults. *Aging, Cognition and Neuropsychology*, *15*, 257–280.

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training and the brain: A review and future directions. *Neuropsychology Review*, *19*, 504–522.

Maddox, G. B., Balota, D. A., Coane, J. H., & Duchek, J. M. (2011). The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. *Psychology and Aging*, *26*, 661–670.

McDaniel, M. A., & Masson, M. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 371–385.

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*, 142–147.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1170–1187.

Nyberg, L. (2005). Cognitive training in healthy aging: A cognitive neuroscience perspective. In R. Cabeza, L. Nyberg, & D. Parks (Eds.), *Cognitive neuroscience of aging: Linking cognitive and cerebral aging* (pp. 218–245). London: London University Press.

Pecher, D., & Raaijmakers, J. G. W. (2004). Priming for new associations in animacy decision: Evidence for context dependency. *Quarterly Journal of Experimental Psychology*, *57*, 1211–1231.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335.

Rabinowitz, J. C., & Craik, F. I. M. (1986). Prior retrieval effects in young and old adult. *Journal of Gerontology*, *41*, 368–375.

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology*, *36*, 325–344.

Rebok, G. W., Carlson, M. C., & Langbaum, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: Traditional and novel approaches. *Journals of Gerontology: Series B*, *62*, 53–61.

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damag. *Perceptual and Motor Skills*, *8*, 271–276.

Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies, & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Brighton, UK: Psychology Press.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced Learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre, & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford: Elsevier.

Rogers, W. A., & Gilbert, D. K. (1997). Do performance strategies mediate age-related deficits in associative learning? *Psychology and Aging*, *12*, 620–633.

Scogin, F., & Bienias, J. L. (1988). A three-year follow-up of older adult participants in a memory-skills training program. *Psychology and Aging*, *3*, 334–337.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, *9*, 371–377.

Troyer, A. K. (2001). Improving memory knowledge, satisfaction, and functioning via an education and intervention program for older adults. *Aging Neuropsychology and Cognition*, *8*, 256–268.

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.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *11*, 164–178.

Wechsler, D. (Ed.). (1997). *Wechsler Adult Intelligence Scale—3rd ed. (WAIS-3®)*. San Antonio, TX: Harcourt Assessment.

West, R. L., Welch, D. C., & Yassuda, M. S. (2000). Innovative approaches to memory training for older adults. In R. D. Hill, L. Backman, & A. S. Neely (Eds.), *Cognitive rehabilitation in old age* (pp. 81–105). New York: Oxford University Press.

Please cite this article in press as: Coane, J.H. Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition* (2013), http://dx.doi.org/10.1016/j.jarmac.2013.04.001