Psychology and Aging

Testing Can Enhance Episodic Memory Updating in Younger and Older Adults

Paige L. Kemp, Vanessa M. Loaiza, and Christopher N. Wahlheim

Online First Publication, September 4, 2023. https://dx.doi.org/10.1037/pag0000776

CITATION
Kemp, P. L., Loaiza, V. M., & Wahlheim, C. N. (2023, September 4). Testing Can Enhance Episodic Memory Updating in Younger and Older Adults. *Psychology and Aging*. Advance online publication. https://dx.doi.org/10.1037/pag0000776

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology and Aging

ISSN: 0882-7974

https://doi.org/10.1037/pag0000776

Testing Can Enhance Episodic Memory Updating in Younger and Older Adults

Paige L. Kemp[1], Vanessa M. Loaiza[2], and Christopher N. Wahlheim[1]
[1] Department of Psychology, University of North Carolina at Greensboro
[2] Department of Psychology, University of Essex

Older adults sometimes show impaired memory for recent episodes, especially those that are similar but not identical to existing memories. Two experiments examined if interpolated testing between episodes improves recent memories for older and younger adults ($N = 60$ per group and experiment). Participants studied two lists of cue–response word pairs. Some pairs included the same cue in both lists with changed responses. Between lists, List 1 pairs were tested (Experiments 1 and 2), tested with corrective feedback (Experiment 1 only), or restudied (Experiment 2 only). On a final cued recall test, participants attempted to recall the List 2 response, indicated if responses had changed between lists, and if so, attempted to recall the List 1 response. List 2 recalls for changed pairs operationalized episodic memory updating. Older adults showed poorer List 2 recall than younger adults. But both age groups showed improved List 2 recall following interpolated testing with or without feedback compared to no-test and restudy contrast conditions. This so-called forward testing effect was accompanied by improved memory for responses having changed across lists. These results contrast with the inhibitory deficit proposal that older adults should be more interference prone than younger adults when competing responses are more accessible during encoding. These findings are more compatible with the view that retrieval practice of competing responses can support the encoding of cross-episode associations and potentially mitigate interference, thus improving age-related associative memory deficits.

Public Significance Statement

This study examined if a retrieval practice technique that can improve younger adults' memory for recent episodes also benefits older adults. By promoting retrieval of existing memories, testing prior information before new learning improved younger and older adults' memory for more recent episodes. These findings suggest that retrieving existing memories can promote memory for more recent related information, thus reducing the mental clutter that sometimes impedes older adults' memory accuracy.

Keywords: associative memory, interference, retrieval practice, proactive effects, testing effect

Supplemental materials: https://doi.org/10.1037/pag0000776.supp

Everyday life is filled with changes that require people to update their memories to guide their future behaviors. People you know may change their last names when they get married, their dietary choices when they receive health news, or beliefs on controversial issues when they learn contradictory information. Remembering the most recent details requires updating memory to prioritize currently relevant information. But outdated long-term memories can proactively interfere with remembering changes. This can have outsized effects on older adults who show age-related deficits in many episodic memory tasks (Balota et al., 2000; Park & Festini, 2017; Salthouse, 2011) and are generally more susceptible to interference (Hasher & Zacks, 1988; Jacoby et al., 2001). This age-related difference signals the need to identify mnemonic techniques that can reduce interference and enhance episodic memory updating.

Theoretical views of age-related episodic memory differences may guide the selection of such techniques. One view proposes that older adults have inhibitory deficits that lead to mental clutter that maintains inappropriate associations among target and competing memories (Amer et al., 2022; Hasher & Zacks, 1988). Accordingly, older adults should experience more interference than younger adults when searching for target memories and attempting to suppress competing memories (e.g., Lustig et al., 2007). However, this view has been updated to account for findings suggesting that excessive links among memories, or hyperbinding, can sometimes support retrieval accuracy by enriching the content of the representations (e.g., Biss et al., 2013). Another view proposes that older adults' recollection deficit (e.g., Hay & Jacoby, 1999; Jacoby, 1999) undermines the extent that

---

Paige L. Kemp https://orcid.org/0000-0002-2200-7735

This research was completed in partial fulfillment of a master's degree by Paige L. Kemp and was presented as a poster at the 2022 Cognitive Aging Conference, Atlanta, Georgia, United States. The stimuli, deidentified data, and analysis scripts are available on the Open Science Framework at https://osf.io/bxhwa/.

Correspondence concerning this article should be addressed to Paige L. Kemp, Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart Building, P.O. Box 26170, Greensboro, NC 27402-6170, United States. Email: plkemp@uncg.edu

event details cue retrieval of existing memories and associated contextual information. This in turn presents fewer opportunities to create associations across episodes that can enrich memories by incorporating details such as the temporal relationship between events (e.g., Wahlheim, 2014). Older adults should thus recollect fewer details that distinguish target from competing memories, such as the original sources.

Despite proposing different mechanisms, the accounts above lead to one prediction: Existing memories should enrich recent episodic memories for older and younger adults when the associations include distinguishing information. In contrast, older adults should experience more interference when memories with shared features are not properly linked. A mnemonic technique that may therefore repair age-related episodic memory updating deficits is testing existing memories before presenting new information with similar but not identical features. By promoting the act of retrieval, testing has been shown to enhance memory for new information and counteract proactive interference from outdated information (for reviews, see Pastötter & Bäuml, 2014; Yang et al., 2018). This *forward testing effect* has been observed across stimuli and testing formats (for a meta-analysis, see Chan et al., 2018). Despite its efficacy, it has rarely been used in the study of cognitive aging (cf. Pastötter & Bäuml, 2019). The present study fills this gap by examining if older and younger adults' memory for new information is comparably improved by prior retrieval of existing memories with shared and novel features.

Interpolating tests after study trials have consistently been shown to reduce prior-list interference and improve subsequent recall (e.g., Divis & Benjamin, 2014; Szpunar et al., 2008). However, only one published study used this multitrial free recall paradigm to examine the forward testing effect in adults older than typical college-aged students. That study used a cross-sectional approach to characterize the forward testing effect in four age groups (Pastötter & Bäuml, 2019). Participants in their 40s, 50s, 60s, and 70s studied three lists, expecting a final recall test. After each of the first two lists, participants either completed a free recall test or restudied the prior list. After the third list, all participants completed a free recall test of that list. All age groups showed a forward testing effect: Final list recall was better when the task interpolated among study trials was a free recall test rather than restudying the prior list.

The consistent forward testing effects in multitrial free recall across adult age groups also suggest that other instances of forward testing effects in younger adults could generalize to older adults. However, the multitrial free recall studies, including lists of individually presented words, do not indicate if such effects will occur when existing memories share features with new information. Because the stimuli in those tasks shared little meaning across lists, testing likely enhanced new learning through some combination of differentiating list contexts (Bäuml & Kliegl, 2013; Szpunar et al., 2008), upregulating attention at the start of new lists (Pastötter et al., 2011), and inducing test expectancy that led to more effective encoding strategies (Weinstein et al., 2011). From a strict inhibitory deficit view, testing in those free recall tasks benefitted older adults by reducing cross-episode associations and attendant mental clutter (cf. Amer et al., 2022).

Findings from younger adult studies suggest that retrieval-induced differentiation could benefit older adults' memory for new information that shares features with existing memories. This may occur, for example, in A-B, A-D paradigms that include pairs with cues that repeat (A) and responses that change (B → D) across lists. Prior studies using this task have shown that retrieving B responses in an interpolated cued recall phase before studying A-D pairs can reduce interference on a subsequent cued recall test (e.g., Malis, 1970; Tulving & Watkins, 1974). Those findings were initially interpreted as implicating a list differentiation mechanism of interference reduction (e.g., Tulving & Watkins, 1974). However, there is mounting evidence that repeating features across episodes, such as the A cue term, lead to remindings of prior-list items that promote cross-episode (list) associations (for a review, see Wahlheim et al., 2021). It is therefore likely that when new information shares features with existing memories, older adults will show forward testing effects that depend on prior-list retrievals that enable enriched (Amer et al., 2022) or integrative (Wahlheim, 2014) encoding of existing and recent memories.

The proposal that episodic memory retrieval can enhance new learning by promoting associative encoding is supported by related studies using A-B, A-D paradigms. Those studies inferred such associations from the degree of dependence in recall of changed responses (D) when earlier responses (B) were also recalled (cf. Bellezza & Schirmann, 1975). One task variant manipulated whether B responses were tested during A-D learning (Postman & Gray, 1977). Subsequent recall of D responses, both overall and conditioned on recall of B responses, was both higher when B responses were retrieved during A-D learning. Similar conditional recall was observed in later studies where, during A-D study, participants indicated if pairs had changed from the prior list and, if so, attempted to recall B responses (for a review, see Wahlheim et al., 2021). Detecting changes during study, which involves overt retrieval of existing memories, has been shown to be associated with better memory for new information when earlier information is also recalled (for related findings in A-B, A-D tasks, see Davis & Chan, 2015; Finn & Roediger, 2013). Studies of age differences in the associations between retrieval of earlier and more recent information have shown weaker associations for older than younger adults (Stawarczyk et al., 2020; Wahlheim, 2014; Wahlheim & Zacks, 2019; but see Garlitch & Wahlheim, 2021). These findings suggest that older adults bind details across events less effectively than younger adults, which is a kind of associative encoding deficit (cf. Naveh-Benjamin, 2000).

The findings from A-B, A-D paradigms summarized above suggest that older adults' impaired updating of episodic memory could be improved by increasing access to existing memories before presenting related information. This would promote enriched, integrative memory representations that support subsequent memory for recent information and its relationship to existing memories. This may be accomplished by testing A-B pairs between lists to increase their accessibility when the shared cues of A-D pairs trigger remindings of A-B pairs, as suggested by an earlier study of younger adults (Wahlheim, 2015). Participants in those experiments studied two lists, including A-B, A-D word pairs, for which between lists, some A-B pairs were tested while others were restudied. During final cued recall, participants recalled List 2 (D) responses and indicated if responses had changed between lists. Participants did not overtly recall List 1 (B) responses, but earlier findings showed that remembering changes is often based on such recalls (Wahlheim & Jacoby, 2013). Successful retrieval during interpolated testing counteracted proactive interference and increased recollection that changes had occurred. Recollecting change was associated with proactive facilitation in List 2 recall, suggesting that testing benefitted new learning partly by promoting cross-episode associations between responses. Similar facilitation should therefore be observed for older adults in the present study to the extent that they successfully retrieve B responses during interpolated testing before studying A-D pairs.

The Present Study

In two experiments using A-B, A-D paradigms, we tested the hypothesis that interpolated testing will improve episodic memory updating for older and younger adults. We examined forward testing effects by manipulating the tasks interpolated between lists for A-B, A-D items. Both experiments included interpolated test without feedback and no-test contrast conditions. The experiments also included contrast conditions to assess the effects of re-exposure to A-B responses by including feedback after interpolated tests (Experiment 1) and restudy of complete A-B pairs (Experiment 2). On final cued recall trials, participants attempted to recall the List 2 response, indicated if the response had changed, and if so, attempted to recall the List 1 response. This sequence allowed us to examine recall dependencies from which to infer successful encoding of cross-episode associations (for a review, see Wahlheim et al., 2021). To rule out a strict inhibitory deficit account, we selected experimental stimuli with overlapping orthography for A-B and A-D responses that had produced no overall proactive interference and sometimes proactive facilitation effects in earlier studies (e.g., Jacoby et al., 2015; Wahlheim & Jacoby, 2013). We assumed that the feature overlap would create more salient retrieval cues, thus triggering A-B remindings during A-D study and mitigating interference, consistent with an integrative encoding view. We thus expected that younger and older adults would both show greater overall proactive facilitation following interpolated testing.

Experiment 1

Experiment 1 was the first to examine forward testing effects on episodic memory updating in older and younger adults in an A-B, A-D task. We compared three A-B, A-D conditions with varying interpolated tasks: tests without feedback, tests with feedback, and no-test contrast trials. The proactive effects of A-B exposure and interpolated tasks were assessed by comparing final cued recall for the A-B, A-D conditions with recall in the C-D control condition, which included pairs that only appeared in the second study list. We expected older adults to show lower overall memory performance across conditions based on established age-related memory differences (for a review, see Rhodes et al., 2019). However, it was unclear if this age-related difference would be comparable across conditions due to mixed prior findings in similar paradigms. Some studies have shown greater age-related differences in recall for A-B, A-D than control pairs (e.g., Ebert & Anderson, 2009; Jacoby et al., 2001; Wahlheim, 2014; Wahlheim & Zacks, 2019), whereas others reported uniform deficits (e.g., Garlitch & Wahlheim, 2020, 2021).

Our primary hypotheses concerned the effects of interpolated testing of A-B pairs on subsequent memory. We attempted to equate interpolated retrieval success for older and younger adults by including test prompts with complete cues and response fragments. We also included the interpolated testing with feedback condition to determine if the forward testing effect was comparable for older and younger adults when they were re-exposed to B responses after all retrievals. Because we expected high interpolated test accuracy given the provision of response fragments, we expected both age groups to show *proactive facilitation* in the form of better List 2 recall for A-B, A-D pairs in the interpolated test conditions than C-D control condition. We also expected interpolated tests to increase memory for the fact that A-B, A-D pairs had changed and recall of B responses by promoting detection of changes during List 2 study. Finally, we expected List 2 recall of D responses to be substantially higher when B responses from List 1 were also recalled, thus reflecting cross-episode associations formed when A-D pairs reminded participants of A-B pairs from List 1. Consistent with this view, to the extent that older adults recall fewer D responses than younger adults, there should be attendant age differences in memory for changes and B responses.

Method

Transparency and Openness

We report how we determined sample sizes, all data exclusions, all manipulations, and all measures. The deidentified data on which the study conclusions are based, the analytical code necessary to reproduce analyses, and the materials used in this study are freely available on the Open Science Framework (OSF) at https://osf.io/bxhwa/ (Kemp et al., 2023). We did not preregister this study. The research reported here was approved by the institutional review board at the University of North Carolina at Greensboro (UNCG) under the title Interpolated Testing and Proactive Effects (No. 20-0073).

Participants

The stopping rule was to collect usable data from 60 younger and 60 older adults based on our available resources and prior work from our lab showing that these sample sizes were sufficient to detect age-related differences in a recall-based episodic memory task (Wahlheim & Garlitch, 2020). These sample sizes also allowed us to test an equal number of participants in each of the five experimental formats. Data collection took place from September 2019 to September 2021 in an urban area in North Carolina, United States. The final sample included 60 younger adults (44 women, 15 men, one gender diverse) ages 18–26 (M = 18.73, SD = 1.36) from UNCG, and 60 older adults (38 women, 22 men) ages 65–84 (M = 71.02, SD = 4.96) from Greensboro and the surrounding areas. The younger adult sample comprised: 35% African American (21), 33% Caucasian (20), 12% Hispanic or Latino (seven), 12% multiracial (seven), 5% Asian or Pacific Islander (three), and 3% other (two) participants. The older adult sample comprised: 80% Caucasian (48), 12% African American (eight), 2% Hispanic or Latino (one), 2% multiracial (one), 2% Asian or Pacific Islander (one), and 2% American Indian and Native American (one) participants. Two younger adults were replaced due to experimenter error, and one older adult was replaced for scoring below the cutoff on the MMSE (62 younger and 61 older adults were tested). For compensation, younger adults received course credit, and older adults received $10 per hour.

We assessed older adults' cognitive health over the phone with the Short Blessed Test (SBT; Katzman et al., 1983) and then in person with the Mini-Mental State Examination (MMSE; Folstein et al., 1975). Older adults had a weighted SBT error score ≤ 4, an MMSE score ≥ 25, and a score of 20/50 or better with one or both eyes on the Snellen eye test of visual acuity (Hetherington, 1954). Table 1 displays years of education and performance on cognitive tasks. Compared to younger adults, older adults scored higher on the Shipley Institute of Living Vocabulary subtest (Shipley, 1986), $t(116.11) = 12.55$, $p < .001$, and had more education, $t(95.95) = 13.96$, $p < .001$. Younger adults scored higher than older adults on the Digit Symbol Substitution Task (DSST; Wechsler, 1981) in number of symbols copied within 90 s, $t(114.92) = 6.04$, $p < .001$, and number of symbols recalled out of nine, $t(110.32) = 3.38$, $p < .001$. There was no age difference in MMSE scores, $t(113.16) = .93$, $p = .35$.

Design

We used a 2 (age: younger, older) × 5 (item type: A-B, A-B;—, C-D [control], A-B, A-D not tested; A-B, A-D tested; A-B, A-D tested with feedback [tested$^{\text{FB}}$]) design. Note that the A-B, A-B item type included filler items that were not of theoretical interest. Age was treated as a between-participants variable, and item type was manipulated within participants.

Materials

The materials were 80 three-word sets taken from Jacoby (1996) that each contained a cue word (e.g., knee) and two response words (e.g., bone, bend) that were orthographically related. The sets were designed so that the two responses could complete the same fragment (e.g., b_n_). The average cue–response forward association strength (Nelson et al., 1998) was .07 ($SD = .10$, range = 0–.49). Seventy-five sets included critical items. Five sets included buffers. Five groups of 15 critical items were equated on the word frequency of cues ($M = 9.36$, $SD = 1.79$, range = 5–14) and responses ($M = 9.35$, $SD = 1.68$, range = 4–14) according to Hyperspace Analogue to Language (HAL) log frequency counts (Lund & Burgess, 1996) taken from the English Lexicon Project database (Balota et al., 2007). The groups were also equated on the length of cues ($M = 5.40$, $SD = 1.58$, range = 3–9 letters) and responses ($M = 4.72$, $SD = 1.21$, range = 3–8 letters). Five experimental formats resulted from rotating groups through conditions. The groups appeared equally often in each condition across participants. The buffer items (that remained constant across formats) appeared (1) at the beginning and end of both study lists to control for primacy and recency effects for critical items, (2) as fillers intermixed among critical items in the interpolated test phase, and (3) as practice items on the final test.

The experiment included four phases: List 1 study, interpolated task, List 2 study, and final cued recall (see Figure 1, for a schematic of the procedure). List 1 comprised 64 word pairs (60 critical, four buffers [two primacy; two recency]) divided evenly across each of four groups that later became A-B, A-B, or A-B, A-D items. The A-B, A-D items were split among interpolated test (tested), interpolated test with feedback (tested$^{\text{FB}}$), and no interpolated test (not tested) conditions. The interpolated task comprised 32 word pairs (30 critical, two fillers) from List 1 divided into groups as just described. List 2 included 80 word pairs (75 critical, five buffers [two primacy; three recency]) evenly distributed across all conditions (15 critical and one buffer each). The A-B, A-B filler items repeated exactly across lists, and the A-B, A-D items included cues (A) that repeated and responses that changed (B → D) across lists. Control items (C-D) only appeared in List 2. The final cued recall test began with five practice trials, including buffer item cues, and then moved to the actual test trials that included all 75 critical item cues.

Procedure

Participants were tested individually. The presentation software (E-Prime 2; Psychology Software Tools, 2012) displayed all stimuli in white font on a black background, except for interpolated test feedback that appeared in green. In all phases, stimuli appeared in a fixed random order with the constraint that no more than three items from the same condition appeared consecutively. The average serial position was equated across conditions.

Table 1

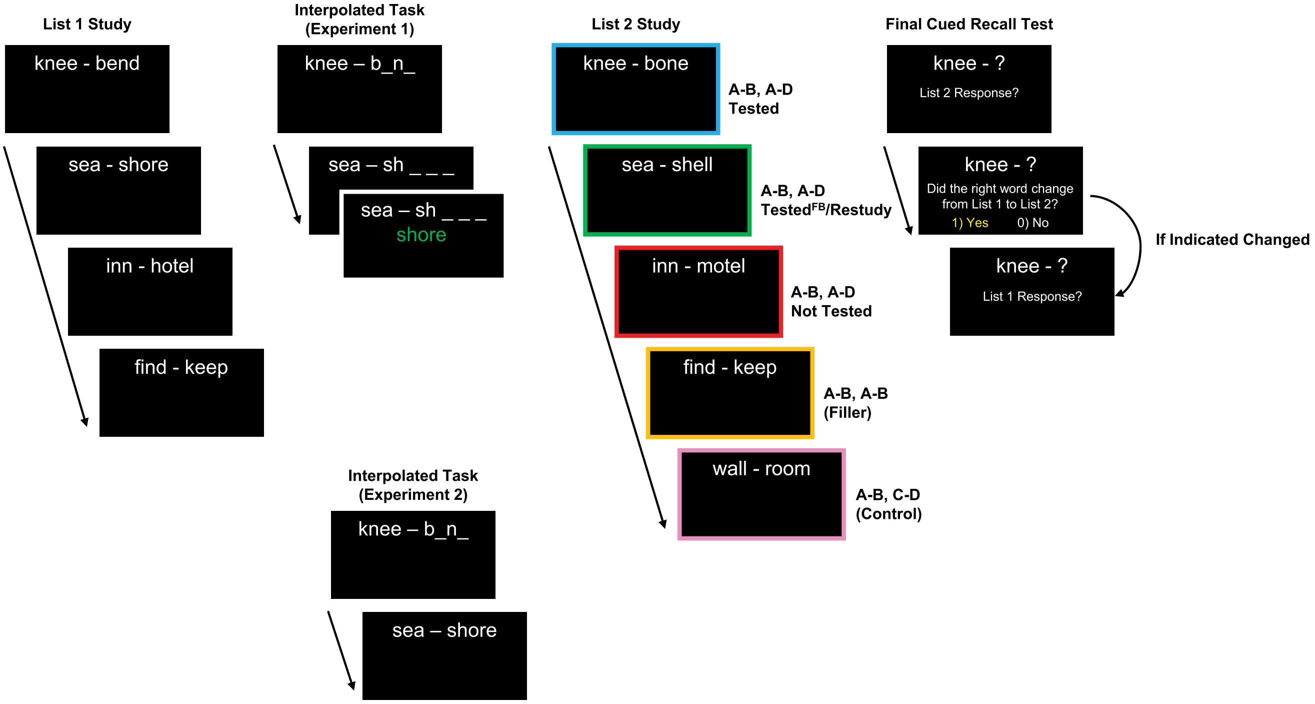
Descriptive Statistics for Education and Performance on Cognitive Tasks: Experiments 1 and 2

		Age			
		Younger		Older	
Experiment	Measure	*M* (*SD*)	Range	*M* (*SD*)	Range
Experiment 1	Education (years)	12.72 (1.15)	12–17	16.78 (1.94)	12–19
	Vocabulary (out of 40)	26.83 (4.12)	17–36	35.72 (3.61)	23–40
	DSST (in 90 s)	68.40 (9.82)	42–87	56.57 (11.58)	33–79
	DSST (out of nine)	7.52 (1.90)	2–9	6.15 (2.49)	1–9
	MMSE	28.38 (1.51)	25–30	28.62 (1.22)	25–30
	SBT (error score)	—	—	0.53 (1.03)	0–4
Experiment 2	Education (years)	12.78 (1.39)	12–19	16.13 (2.05)	12–19
	Vocabulary (out of 40)	27.07 (4.31)	17–37	35.35 (2.87)	28–40
	DSST (in 90 s)	60.25 (9.08)	38–83	49.33 (12.87)	19–77
	DSST (out of nine)	7.50 (1.93)	2–9	5.72 (2.51)	0–9
	MMSE	28.28 (1.33)	25–30	28.35 (1.38)	25–30
	SBT (error score)	—	—	1.25 (1.49)	0–4

Note. Vocabulary = Shipley Institute of Living Scale Vocabulary (Shipley, 1986); DSST = Digit Symbol Substitution Task (Wechsler, 1981); MMSE = Mini-Mental State Examination (Folstein et al., 1975); SBT = Short Blessed Test (Katzman et al., 1983).

Figure 1

Schematic of the Experiment Procedure: Experiments 1 and 2



Note. FB = feedback. A schematic overview of the trial structures from the procedures in both experiments. In both experiments, during the study of List 1, participants read out loud a list of word pairs. The main difference between experiments was the trial structure during the interpolated task: In Experiment 1, participants were provided a cue along with a fragmented response and were asked to recall some of the List 1 responses (A-B, A-D tested). For half of the trials, participants were provided with corrective feedback following their response (A-B, A-D tested$^{\mathrm{FB}}$); in Experiment 2, participants were provided a cue along with a fragmented response and were asked to recall some of the List 1 responses (A-B, A-D tested) or they were reshown pairs from List 1 and asked to restudy them (A-B, A-D restudied). In both experiments, during the study of List 2, participants read out loud a list of word pairs that had the same cue as List 1 but a changed response (blue, green, and red borders) that was repeated across Lists (orange border), or were new (pink border). During the final cued recall test, participants first recalled List 2 responses, then indicated if the response changed across lists, and for those, attempted to recall the List 1 response. See the online article for the color version of this figure.

During List 1 study, participants read word pairs aloud and studied them for an upcoming test. Pairs appeared for 5,000 ms with a 500 ms interstimulus interval (ISI). During the interpolated phase, participants completed test trials that sometimes included feedback. When cue–fragment pairs (e.g., knee—b_n_) appeared, participants read the cue and recalled List 1 response aloud. All pairs appeared for 7,000 ms each (500 ms ISI), but in the feedback condition, pairs appeared alone for the first 5,000 ms, and then List 1 responses appeared below for 2,000 ms. Participants attempted to respond before the feedback, read the feedback silently, and were discouraged from guessing. An experimenter recorded the responses. During List 2 study, participants read word pairs aloud and studied them for an upcoming test. Pairs appeared for 5,000 ms each (500 ms ISI). They were told that the list would contain List 1 pairs, pairs with List 1 cues and changed responses, and new pairs. Participants were told to silently note when pairs included changes.

During the final cued recall test, each cue appeared with a question mark (e.g., knee—?). Participants recalled the List 2 response (e.g., bone) by typing it onto the screen and indicated if the response had changed by pressing the "1" key for "yes" and the "0" key for "no." When they responded "yes," they were prompted to recall the List 1 response by typing it onto the screen when they responded "no," the program advanced to the next trial. Participants were allowed to pass. After the test, participants completed a computerized Shipley Vocabulary test (Shipley, 1986), and written versions of the MMSE (Folstein et al., 1975) and DSST (Wechsler, 1981).

Statistical Methods

R software (R Core Team, 2021) was used for all statistical tests. To evaluate experimental effects, we fitted logistic mixed-effects models using the *glmer* function from the *lme4* package (Bates et al., 2015). The models included random intercept effects for participants and items, as well as fixed effects of age and item type. Wald's χ^2 hypothesis tests were performed using the analysis of variance function of the *car* package (Fox & Weisberg, 2019). Post hoc pairwise comparisons using the Tukey method were performed using the *emmeans* function of the *emmeans* package (Lenth, 2021). The significance level was $\alpha = .05$.

We also conducted analogous Bayesian logistic mixed-effects models using the *brms* package (Bürkner, 2018). The purpose of these analyses was to compare the relative likelihood of the data under one model (e.g., an alternative model that assumes a main effect of age, M_1) against that of another (e.g., a null model that only includes random effects of participant and item type, M_0). The ratio of these likelihoods is the Bayes factor (BF), which reflects the relative evidence for one model over another and can be interpreted continuously. For example, a BF_{10} of 0.1 suggests that the null model in the denominator is strongly preferred to the alternative model in the numerator by a factor of 10, whereas BF_{10}s ranging from 1 to 3 and 3 to 10, respectively, indicate anecdotal and substantial evidence in favor of the alternative model in the numerator. *BF*s were estimated by comparing models that included a predictor (e.g., a main effect of age) against a model that excluded it (e.g., the null model) using the *bridge sampler* function (Gronau et al., 2020) in the brms package. We used Wetzels et al.'s (2012) method for reporting the *BF*s for the unique effect of each factor and the interaction. Following recommendations of prior work using similar models (Bartsch & Oberauer, 2023; Oberauer, 2019), we applied Cauchy priors (with a scale of 0.353) to the regression coefficients and noninformative Lewandowski–Kurowicka–Joe priors to the correlation matrices (with shape Parameter 1). Each model was fit with four chains of 50,000 iterations each, with the first 1,000 warmup iterations excluded from analysis. We checked for convergence of the chains via visual inspection and verified that the R-hat statistic was close to one for all the parameters of the fitted models. Posterior predictive checks also ensured appropriate model fits to the data.

Results

Interpolated Tests

Interpolated test accuracy (Table 2, left columns) was comparable for both age groups as indicated by no significant effects from an Age × Item Type model, *largest* $\chi^2(1) = 2.60$, $p = .11$, all BF_{10}s < 0.010.

Final Cued Recall Test

List 2 Recall. We examined interpolated testing effects on episodic memory updating by comparing List 2 recall across all conditions except the A-B, A-B fillers (Figure 2A). An Age × Item Type model indicated a significant effect of age, $\chi^2(1) = 13.79$, $p = .001$, $BF = 3.867 \times 10^{25}$, showing higher recall for younger than older adults. There was also a significant effect of item type, $\chi^2(3) = 132.66$, $p < .001$, $BF = 67.396$, showing several differences. Recall was higher in the tested and tested^{FB} conditions than the not tested and control conditions, *smallest* z ratio = 7.09, $p < .001$, showing forward testing benefits to episodic memory updating. However, recall did not differ between the tested and tested^{FB} conditions, z ratio = 0.68, $p = .91$, nor it differ between the not tested and control conditions, z ratio = 1.38, $p = .51$. Finally, the interaction was not significant, $\chi^2(3) = 2.51$, $p = .47$, $BF = 0.075$. These results show that interpolated testing led to high rates of retrieval success that promoted comparable episodic memory updating for older and younger adults.

Table 2
Interpolated Test Accuracy for A-B, A-D Item Types: Experiments 1 and 2

Age	Experiment 1		Experiment 2
	Test	Test^{FB}	Test
Younger	.90 [.86, .93]	.89 [.84, .92]	.89 [.85, .92]
Older	.89 [.84, .92]	.86 [.82, .90]	.90 [.86, .93]

Note. 95% confidence intervals are displayed in brackets. FB = feedback.

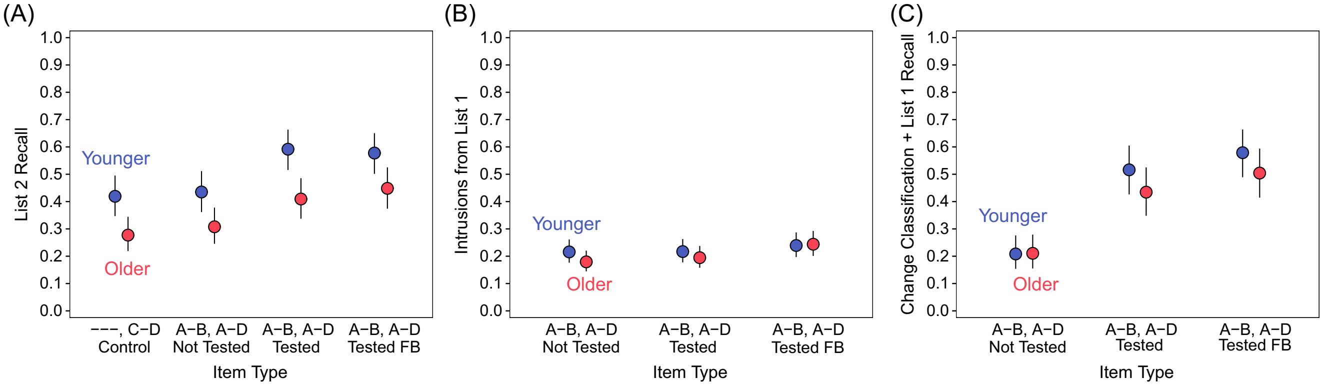
Intrusions From List 1. We further examined interpolated testing effects by comparing intrusion errors from List 1 in the A-B, A-D conditions (Figure 2B). An Age × Item Type model indicated no significant effect of age, $\chi^2(1) = 0.74$, $p = .39$, $BF = 0.353$, a significant effect of item type, $\chi^2(2) = 11.50$, $p < .01$, $BF = 5.113$, and no significant interaction, $\chi^2(2) = 2.56$, $p = .28$, $BF = 0.157$. Pairwise comparisons showed no significant differences between the not tested and tested conditions, z ratio = 0.65, $p = .79$. But there were significantly more intrusions in the tested^{FB} than not tested and tested conditions, *smallest* z ratio = 2.57, $p = .03$, indicating that when alternative responses were presented as feedback, this created more proactive interference.

Change Classifications. If interpolated testing benefitted List 2 recall by promoting cross-episode associations during List 2, then more changes and List 1 responses should be remembered in the interpolated test conditions. We assessed this by examining differences in *change recollection*, which includes A-B, A-D items classified as changed and accurate List 1 recall (Figure 2C). We used this measure to assay differences in the extent that interpolated testing presented opportunities for cross-episode binding. Indeed, manipulations that affect in List 1 recall during List 2 study have consistently shown parallel effects on change recollection (for a review, see Wahlheim et al., 2021). For completeness, we also report these rates in Supplemental Materials (henceforth Supplemental Material) Table S1 alongside the rates for A-B, A-D items classified as changed with inaccurate List 1 recall (*change remembered, not recollected*) and not classified as changed (*change not remembered*). An Age × Item Type model of change recollection, including only the A-B, A-D conditions, indicated no significant age effect, $\chi^2(1) = 1.06$, $p = .30$, $BF = 0.558$, a significant item type effect, $\chi^2(2) = 352.69$, $p < .001$, $BF = 4.572 \times 10^{80}$, and no significant interaction, $\chi^2(2) = 5.17$, $p = .08$, $BF = 0.852$. Change recollection was higher in the test and tested^{FB} conditions than not tested condition, *smallest* z ratio = 14.84, $p < .001$, and was higher for the tested^{FB} than test condition, z ratio = 3.46, $p < .01$. Collectively, these findings suggest that interpolated testing presented more opportunities to encode cross-episode associations, especially when feedback was provided.

List 2 Recall Conditionalized on Change Classifications. Many studies have shown that List 2 recall is especially accurate when changes are recollected (for a review, see Wahlheim et al., 2021). The combination of facilitated recall with more recollected changes following manipulations that increase List 1 recall during List 2 indexes improvements in cross-episode binding. We followed this logic for assessing integrative encoding differences by conditionalizing List 2 recall on recollected changes (Figure 3, left panel). An Age × Item Type model including only the A-B, A-D conditions indicated a significant age effect, $\chi^2(1) = 32.82$, $p < .001$, $BF = 65{,}997.063$, showing higher recall for younger than older adults. The item type and interaction effects were not significant, *largest*, $\chi^2(2) = 4.85$, $p = .09$, *BF*s ≤ 0.676. These results suggest that younger adults associated responses across lists more successfully during encoding than older adults.

Figure 2

Correct List 2 Recall, Intrusions From List 1, and Change Classification and List 1 Recall on the Final Cued Recall Test: Experiment 1



Note. FB = feedback. Probabilities of List 2 recall (Panel A), intrusions from List 1 (Panel B), and change classification and List 1 recall (Panel C) on the final test as a function of age and item type in Experiment 1. Points are probabilities estimated from mixed-effects models; error bars are 95% confidence intervals. Blue points represent probabilities for younger adults, and red points represent probabilities for older adults. See the online article for the color version of this figure.

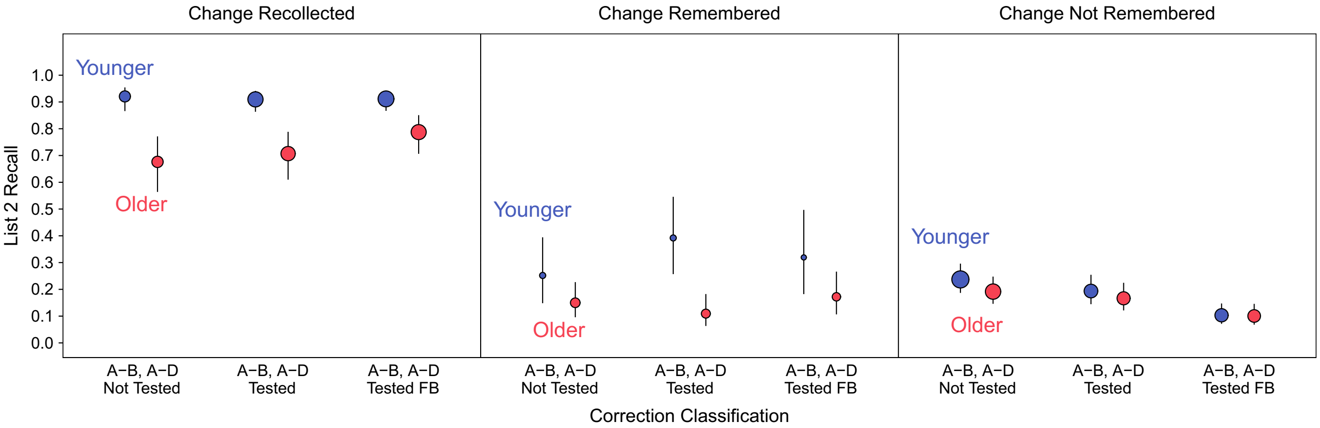
For completeness, we examined List 2 recall conditionalized on the other classifications using the same model as previously for each classification. The model for remembered but not recollected changes (middle panel) indicated a significant age effect, $\chi^2(1) = 14.83$, $p < .001$, $BF = 265.374$, showing higher recall for younger than older adults. No other effects were significant, *largest* $\chi^2(2) = 3.83$, $p = .15$, BFs ≤ 1.361. Additionally, the model for unremembered changes (right panel) indicated a significant item type effect, $\chi^2(2) = 37.65$, $p < .001$, $BF = 3.667 \times 10^6$, showing that recall was significantly lower in the tested$^{\text{FB}}$ condition than the tested and not tested conditions, *smallest* 4.19, $p < .001$. No other effects were significant, *largest* $\chi^2(1) = 2.13$, $p = .14$, BFs ≤ 0.660. These results suggest that providing alternative responses as feedback during interpolated testing created more proactive interference that had its effects when participants were unable to remember changes.

Discussion

Experiment 1 showed that increasing accessibility of List 1 responses via interpolated testing enhanced memory for changed List 2 responses comparably for both age groups, despite older adults recalling less overall. Interpolated test feedback did not lead to memory differences for changed List 2 responses but did increase intrusion errors for both age groups, suggesting that re-exposure to the List 1 responses created proactive interference. There was evidence that the proactive effects of interpolated testing can partially be attributed to improved cross-episode binding. Both age groups

Figure 3

Conditional Correct List 2 Recall on Final Cued Recall Test: Experiment 1



Note. FB = feedback. Probabilities of List 2 recall conditioned on change classifications as a function of age and item type in Experiment 1. Points are probabilities estimated from mixed-effects models; error bars are 95% confidence intervals. Blue points represent probabilities for younger adults, and red points represent probabilities for older adults. Point sizes indicate for each cell the relative proportion of observations, which are also displayed in Supplemental Table S1. See the online article for the color version of this figure.

recollected change more often for A-B, A-D items that were tested in the interpolated phase than those not tested, and change recollection was associated with higher List 2 recall. Providing feedback increased change recollection, but the benefits were offset by greater interference when change was not recollected. Critically, when changes were recollected, younger adults showed higher List 2 recall than older adults, suggesting that an age-related difference in cross-episode binding remained, despite the benefits of interpolated testing.

Experiment 2

Although Experiment 1 provided the first evidence of forward testing effects in older and younger adults in a paired-associate updating task, the contrast condition without interpolated testing did not re-expose List 1 responses. It is therefore unclear whether the forward testing effects observed in Experiment 1 reflect retrieval practice per se. Experiment 2 addressed this issue by including an interpolated restudy task. Testing has been shown to improve memory for retrieved information more than restudying at high levels of retrieval success (for a meta-analytic review, see Rowland, 2014). Because interpolated testing with response fragments in Experiment 1 led to high test accuracy, we hypothesized that retrieval practice benefits for memory updating would be greater than those provided by re-exposure to studied pairs. We also assumed that forward testing effects would reflect more opportunities to encode cross-episode associations; therefore, interpolated testing should promote recollection of change better than interpolated restudy. We had no a priori reason to predict age-related differences in such forward testing effects.

Method

Participants

The stopping rule was to collect usable data from 60 younger and 60 older adults based on our available resources and to match the sample sizes from Experiment 1. Data collection took place from August 2022 to March 2023 in an urban area in North Carolina, United States. The final sample included 60 younger adults (45 women, 14 men, one gender diverse) ages 18–29 ($M = 19.02$, $SD = 1.91$) and 60 older adults (43 women, 17 men) ages 65–89 ($M = 71.78$, $SD = 5.35$) from the same populations as Experiment 1. The younger adult sample comprised 43% Caucasian (26), 33% African American (20), 10% Hispanic or Latino (six), 7% multiracial (four), 5% Asian or Pacific Islander (three), and 2% American Indian and Native American (one) participants. The older adult sample comprised 73% Caucasian (44), 17% African American (10), 3% American Indian and Native American (two), 3% other (two), 2% Hispanic or Latino (one), and 2% multiracial (one) participants. Nine younger adults were replaced because of MMSE scores below cutoff (six), computer malfunction (two), and voluntary withdrawal (one); four older adults were replaced because of computer malfunction (three) and an MMSE score below cutoff (one); one middle-aged adult (39 years of age) was excluded for not fitting into the younger and older adult age ranges. A total of 69 younger, 64 older, and one middle-aged adults were tested. For compensation, younger adults received course credit, and older adults entered a raffle for six Amazon gift cards ($45 each).

Older adults' cognitive health was assessed as in Experiment 1. Table 1 displays years of education and performance on cognitive tasks. Compared to younger adults, older adults scored higher on the Shipley Institute of Living Vocabulary subtest (Shipley, 1986), $t(102.89) = 12.39$, $p < .001$, and had more education, $t(103.72) = 10.46$, $p < .001$. Younger adults scored higher than older adults on the DSST (Wechsler, 1981) in number of symbols copied within 90 s, $t(106.10) = 5.37$, $p < .001$ and number of symbols recalled out of nine, $t(110.58) = 4.36$, $p < .001$. There was no age difference in MMSE scores, $t(117.86) = .27$, $p = 79$.

Design, Materials, and Procedure

We used the Experiment 1 design, replacing the A-B, A-D tested^{FB} condition with an A-B, A-D restudied condition. The materials and procedure were identical to Experiment 1, but the interpolated task comprised an equal number of cue–fragment test pairs and cue–response restudy pairs that were repeated from List 1.

Results

Interpolated Tests

Interpolated test accuracy (Table 2, right panel) was comparable for younger and older adults, $\chi^2(1) = 2.06$, $p = .15$, $BF = 0.008$.

Final Cued Recall Test

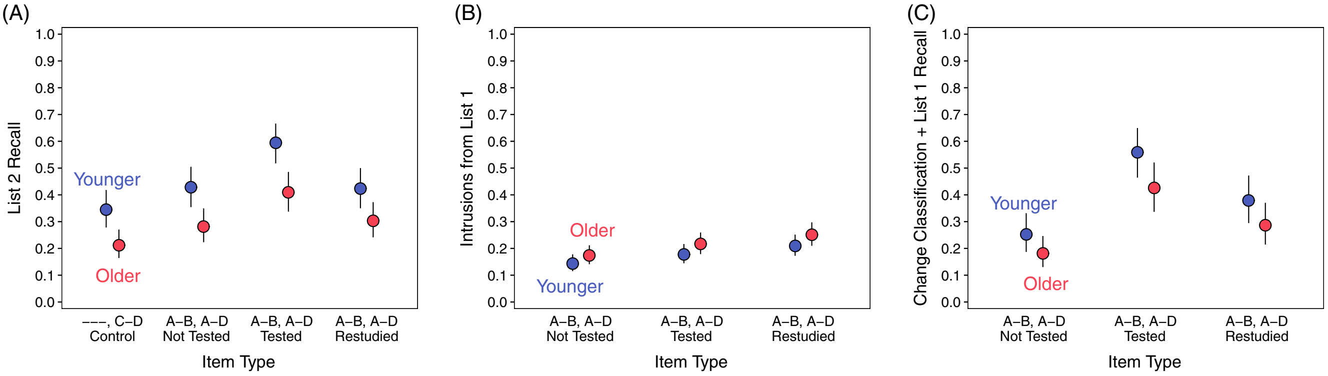
Correct List 2 Recall. As in Experiment 1, we examined interpolated testing effects on episodic memory updating by comparing List 2 recall across all conditions except the A-B, A-B fillers (Figure 4A). An Age × Item Type model indicated a significant age effect, $\chi^2(1) = 14.75$, $p < .001$, $BF = 1.103 \times 10^{33}$, showing higher recall for younger than older adults. There was also a significant item type effect, $\chi^2(3) = 164.25$, $p < .001$, $BF = 92.835$, showing several differences: Recall was higher in all A-B, A-D conditions than in the control condition, *smallest* z ratio = 4.64, $p < .001$. In the A-B, A-D conditions, recall was higher for the tested than restudied and not tested conditions, *smallest* z ratio = 7.62, $p < .001$, and did not differ between the latter conditions, z ratio = 0.54, $p = .95$. Finally, the interaction was not significant, $\chi^2(3) = 2.23$, $p = .53$, $BF = 0.058$. These results replicate the forward testing benefit for older and younger adults even when contrasted with a condition that re-exposed all List 1 responses during the interpolated phase.

Intrusions From List 1. Also as in Experiment 1, we further examined interpolated testing effects by comparing intrusions errors from List 1 in the A-B, A-D conditions (Figure 4B). An Age × Item Type model indicated a significant age effect, $\chi^2(1) = 4.21$, $p = .04$, $BF = 1.380$, showing more intrusions for older than younger adults. There was also a significant item type effect, $\chi^2(2) = 29.81$, $p < .001$, $BF = 22{,}691.711$, showing more intrusions in the tested and restudied conditions than in the not tested condition, *smallest* z ratio = 3.03, $p = .01$, and in the restudied condition than tested condition, z ratio = 2.43, $p = .04$. The interaction was not significant, $\chi^2(2) = 0.02$, $p = .99$, $BF = 0.090$.

Change Classifications. We again examined differences in the potential encoding of cross-episode associations, by assessing change classifications, focusing on change recollection (Figure 4C). We report the other A-B, A-D classification rates in Supplemental Table S2. An Age × Item Type model, including only the A-B, A-D conditions, indicated a

Figure 4

Correct List 2 Recall, Intrusions From List 1, and Change Classification and List 1 Recall on the Final Cued Recall Test: Experiment 2



Note. Probabilities of List 2 recall (A), intrusions from List 1 (B), and change classification and List 1 recall (C) on the final test as a function of age and item type in Experiment 2. Points are probabilities estimated from mixed-effects models; error bars are 95% confidence intervals. Blue points represent probabilities for younger adults, and red points represent probabilities for older adults. See the online article for the color version of this figure.

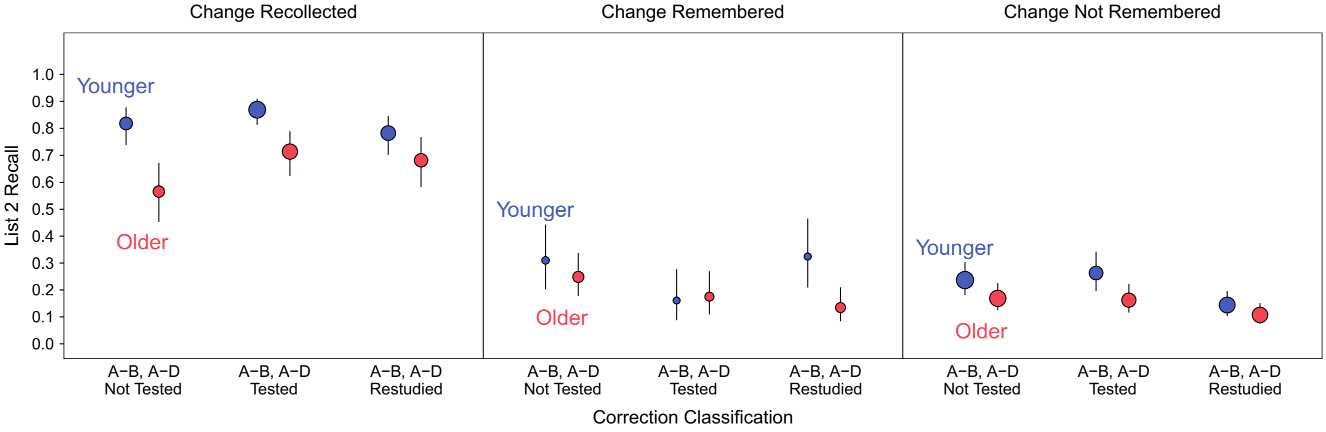
significant age effect, $\chi^2(1) = 4.24$, $p = .04$, $BF = 1.730$, showing higher change recollection for younger than older adults, and a significant item type effect, $\chi^2(2) = 237.39$, $p < .001$, $BF = 4.555 \times 10^{50}$, showing significant differences in change recollection in the following descending order: tested > restudied > not tested, *smallest z* ratio = 7.25, $p < .001$. The interaction was not significant, $\chi^2(2) = 0.67$, $p = .72$, $BF = 0.141$.

List 2 Recall Conditionalized on Change Classifications. We again assessed cross-episode binding differences by conditionalizing List 2 recall on recollected changes (Figure 5, left panel). An Age × Item Type model, including only the A-B, A-D conditions, indicated a significant age effect, $\chi^2(1) = 17.28$, $p < .001$, $BF = 217.755$, showing higher recall for younger than older adults. A significant item type effect, $\chi^2(2) = 14.53$, $p < .001$, $BF = 36.651$, showed higher recall in the tested than restudied and not tested conditions, *smallest z* ratio = 2.82, $p = .01$, and no difference in the latter conditions, *z* ratio = 0.89, $p = .65$. Their interaction was not significant, $\chi^2(2) = 6.01$, $p = .05$, $BF = 0.697$. These results suggest that younger adults associated responses across lists better than older adults; they also suggest that interpolated testing best supported such encoding.

For completeness, we examined List 2 recall conditionalized on the other classifications using the same model as previously for each classification. The model for remembered but not recollected changes (middle panel) indicated no significant age effect, $\chi^2(1) = 3.56$, $p = .06$, $BF = 1.132$, a significant item type effect, $\chi^2(2) = 6.98$, $p = .03$, $BF = 1.369$, and no significant interaction, $\chi^2(2) = 5.64$, $p = .06$, $BF = 1.721$. Recall was higher in the not tested than tested condition, *z* ratio = 2.51, $p = .03$; recall did not differ between these conditions and the restudied condition, *smallest z* ratio = 1.11, $p = .51$. Additionally, the model for unremembered changes (right panel) indicated a significant age

Figure 5

Conditional Correct List 2 Recall on Final Cued Recall Test: Experiment 2



Note. Probabilities of List 2 recall conditioned on change classifications as a function of age and item type in Experiment 2. Points are probabilities estimated from mixed-effects models; error bars are 95% confidence intervals. Blue points represent probabilities for younger adults, and red points represent probabilities for older adults. Point sizes indicate for each cell the relative proportion of observations, which are also displayed in Supplemental Table S2. See the online article for the color version of this figure.

effect, $\chi^2(1) = 7.60$, $p = .01$, $BF = 5.628$, showing higher recall for younger than older adults, and a significant item type effect, $\chi^2(2) = 25.29$, $p < .001$, $BF = 7{,}713.658$, showing lower recall in the restudied than tested and not tested conditions, *smallest* z ratio = 4.33, $p < .001$, and no difference in the latter conditions, z ratio = 0.35, $p = .93$. There was no significant interaction, $\chi^2(2) = 1.01$, $p = .60$, $BF = 0.291$. These results suggest that providing alternative responses via re-exposure during restudied trials created more proactive interference when participants were unable to remember changes.

Discussion

Experiment 2 replicated Experiment 1 in showing that interpolated testing of A-B pairs enhanced memory for changed List 2 responses comparably for both age groups, and older adults recalled less overall. Additionally, interpolated testing led to greater episodic memory updating than interpolated restudy for both age groups, suggesting that the forward testing effects reflect retrieval practice as opposed to re-exposure to the List 1 pairs. In contrast to Experiment 1, interpolated testing of A-B, A-D items led to more intrusion errors of List 1 responses compared to nontested items, and older adults showed high intrusion rates. But as in Experiment 1, there was evidence that interpolated testing enhanced episodic memory updating in part by promoting the cross-episode binding. These findings are sensible given that conditions that increase the accessibility of existing memories also increase the potential for interference, especially for older adults. Both age groups recollected more changes when List 1 responses were tested during the interpolated phase, and younger adults showed higher List 2 recall than older adults when recollecting changes at test. Additionally, for both age groups, change recollection occurred more often for A-B, A-D items that were tested as opposed to restudied, and this was associated with higher List 2 recall. Collectively, these results are generally more compatible with an integration than inhibition account in showing that practicing retrieval of competing responses improved episodic memory updating and conditional recall that indexed cross-episode associations.

General Discussion

Two experiments examined interpolated testing effects on episodic memory updating in older and younger adults using A-B, A-D tasks. Interpolated testing enhanced memory for changed List 2 responses for both age groups, despite older adults recalling less overall. The complete recall patterns suggested that interpolated testing promoted cross-episode binding during encoding, as interpolated testing led to more recollection of changes that were associated with enhanced List 2 recall. These results contribute to the nascent research on aging and the forward testing effect by showing that interpolated testing can counteract proactive interference in older and younger adults when existing memories share features with new information.

Previous work using A-B, A-D tasks has shown that older adults are generally more vulnerable to interference than younger adults (for reviews, see Kane & Hasher, 1995; Kausler, 1994). While age-related memory differences using this paradigm have been observed (e.g., Arenberg, 1973; Wahlheim, 2014; Wahlheim & Zacks, 2019; Winocur & Moscovitch, 1983), such differences are not always found (e.g., Freund & Witte, 1976; Garlitch & Wahlheim, 2020, 2021). These mixed results reflect moderating effects of variables that affect the accessibility of existing memories when studying stimuli with shared and changed features, such as interpolated testing, and relationships among stimulus features, and population characteristics (e.g., cognitive health status, age, education). These variables determine the extent that competing responses are coactivated during encoding, which could lead to memory impairment or improvement.

Inhibition deficit theory proposes that more coactivation should lead to more interference for older adults who suppress irrelevant information less effectively than younger adults (for reviews, see Campbell et al., 2020; Lustig et al., 2007). A strict interpretation of this view leads to the prediction of consistent age-related memory impairments when experimental conditions promote retrieval of existing memories in new learning contexts. However, studies have shown that impaired suppression can improve task performance by promoting meaningful connections (for a review, see Amer et al., 2022). Alternatively, age-related memory differences could reflect older adults being less able to recollect existing memories when studying related information, thus providing fewer opportunities to integrate the past with the present (for a review, see Wahlheim et al., 2021). This shared perspective holds that forming associative links across contexts can promote episodic memory updating, regardless of age. Here, interpolated testing promoted these connections and improved recall of recent information and memory for its relationship with existing memories, which was more consistent with the view that noticing relationships can counteract response competition and interference.

Although both age groups benefitted from interpolated testing via integrative encoding, older adults showed a weaker association between recall of B and D responses. This may reflect older adults' binding deficits (for a meta-analysis, see Old & Naveh-Benjamin, 2008) disrupting associative encoding when responses were coactivated. Additionally, older adults' downstream recollection deficits may have disrupted retrieval of cross-episode associations (Wahlheim, 2014). Here, we showed that successful retrievals of existing memories may be one way to bolster older adults' associative encoding. Based on work showing that older adults can benefit from mnemonic training (e.g., Kirchhoff et al., 2012), the present findings suggest that study-phase retrievals could be used strategically to support older adults' memory updating.

The finding that interpolated testing benefited older adults' episodic memory updating replicates and extends previous work in younger adults (Wahlheim, 2015). One key difference between these studies is that interpolated testing led to proactive facilitation in overall List 2 recall here, but it eliminated proactive interference before. This discrepancy occurred partly because, unlike the prior study, the interpolated tests here included word fragments that ensured high levels of retrieval success. The discrepancy also occurred partly because the present word sets contained responses with high orthographic overlap (e.g., lawn-*grass*; lawn-*green*), whereas the prior study used responses with less overlapping orthography (e.g., coffee-*table*; coffee-*bean*). The former was likely more effective at cueing retrieval of existing memories during new learning. Collectively, these findings suggest that theories of aging and interpolated testing effects must account for interactions of interpolated retrieval success and the potential for stimulus characteristics to cue study-phase retrievals.

Following the assumption that retrieval success is necessary for interpolated testing to promote integrative encoding, our first experiment also examined how interpolated testing influenced episodic memory updating when accompanied by corrective feedback. Although feedback can provide participants with more opportunities for integrative encoding by re-exposing participants to unrecalled responses during new learning, it can also lead to more proactive interference when subsequent retrieval is not recollection based (Wahlheim, 2015; Experiment 2). Consequently, overall recall of changed responses reflects influences of interpolated test accuracy and subsequent recollection of changes. Here, we showed that feedback provided more opportunities to notice changes as shown by higher recall of responses from both lists on the final test. However, feedback also led to more intrusions of List 1 responses and poorer recall of List 2 responses when participants did not recollect changes on the final test (for a similar discussion, see Chan et al., 2018), presumably because the enhanced List 1 accessibility rendered those responses more likely to be automatically retrieved later.

The benefits and costs of interpolated test feedback have implications for predicting when feedback should be delivered to improve episodic memory updating. Although testing promoted episodic memory updating regardless of feedback here, the impact of its provision may be more pronounced at lower levels of interpolated retrieval success. As mentioned earlier, to ensure successful retrieval of earlier responses, we provided cues along with word fragments for the responses. This was also done in part to equate interpolated recall between younger and older adults because age-related episodic memory differences are smaller when more environmental support is provided (Craik, 2017). Consequently, one may wonder how the outcomes associated with feedback differ when retrieval cues provide less environmental support, particularly for older adults. It seems reasonable that older adults would benefit less when interpolated testing requires more self-generation, leading feedback to be corrective on more trials than here. Although feedback may confer greater benefits to older adults, these benefits may also be offset by the costs of recollection failures that would lead to feedback to produce more interference.

The present results are also relevant for disentangling the mechanisms of interpolated testing benefits. In Experiment 2, we examined the possibility that the benefits of interpolated testing via integrative encoding are unique to the act of retrieval practice as opposed to simply restudying the List 1 pairs. Previous work has shown that when retrieval success is high, testing can improve memory for retrieved information more than restudying (for a meta-analytic review, see Rowland, 2014). The high retrieval success conferred by fragments in Experiment 1 led us to predict the observed outcome that interpolated testing would enhance episodic memory updating more than interpolated restudying. Critically, such benefits were accompanied by improved recollection of changes, suggesting that cross-episode associations were better supported by interpolated testing than restudy. These findings join prior work suggesting that the forward testing effect in paired-associate learning can reflect integrative encoding (Wahlheim, 2015).

Limitations and Future Directions

The present study had several limitations that should be acknowledged. First, we inferred the role of integrative encoding in interpolated testing effects from dependence in the recall of responses from both lists on the final tests. This was by design: An overt measure of List 1 recall during List 2 study was not included to prevent reactive encoding strategies and to isolate interpolated testing effects. We also avoided increasing working memory demands because aging is associated with reduced working memory capacity (for a review, see Bopp & Verhaeghen, 2005), and we did not want to overburden older adults with a secondary List 2 task. The concern was that this would disproportionately distract older adults from encoding changed responses. While we did not measure List 1 recall during List 2 directly, prior findings give us confidence that the change recollection measure was sensitive to differences in upstream integration-based encoding. Indeed, prior work has shown that when participants cannot retrieve List 1 responses during List 2, they almost never retrieve List 1 responses at test (Wahlheim & Jacoby, 2013). Thus, retrieval dependencies assumed to support differences in opportunities for integrative encoding are almost always exclusively observed when study-phase retrievals had occurred upstream. Future work could still verify that the downstream change recollection measure is sensitive to study-phase retrieval differences by measuring List 1 recall during List 2 study.

Second, the proactive effects of interpolated testing were observed using a unique material set with orthographically related responses. As described above, such stimulus characteristics should have consequences for episodic memory updating. The consequences of those characteristics could be systematically explored by manipulating the degree of perceptual and conceptual response similarity. Prior work suggests that integrative encoding is promoted when items are related semantically because this stimulates more remindings (e.g., Hintzman & Block, 1973; McKinley & Benjamin, 2020; Tullis et al., 2014). Assuming that semantic relatedness better supports cross-episode binding than perceptual relatedness, older adults may benefit more from interpolated testing when competing responses are semantically associated. Consistent with this proposal, age-related binding deficits can be mitigated by increasing semantic associations among items (Delhaye et al., 2019; Loaiza & Srokova, 2020). However, semantically related words may also lead to more interference when older adults fail to recollect changes, given that those responses would be less distinctive. Interactions between interpolated testing and semantic relatedness could also be inferred across studies using materials that vary in personal relevance or familiarity to participants, such as videos of everyday actions (Wahlheim & Zacks, 2019) and corrections of fake news headlines (Wahlheim et al., 2020). Currently, we can only speculate how varying degrees of semantic relatedness affects the benefits associated with interpolated testing for older adults, thus indicating the need for future studies to better understand the generalizability of this mnemonic technique across different materials.

Third, the interpolated test cues were paired with response fragments. These fragments varied in the number of blanks, and, thus, possible completions (e.g., do_ vs. sh_ _ _). Interference may be greater when there are more response possibilities leading to more response competition, fewer successful retrieval attempts, and poorer episodic memory updating. We explored this possibility by comparing interpolated recall and List 2 recall accuracy conditioned on the number of fragments (Supplemental Table S1). Interpolated recall accuracy did not differ depending on the number of possible fragment completions but there were differences downstream. List 2 recall was highest for the fragments with one blank and there were no differences among fragments with two blanks and fragments with three blanks or greater. These results imply that more response competition for items during the interpolated phase may negatively impact memory for changed information. However, given the preliminary nature of these findings, future research should test this assumption by directly comparing episodic memory updating as a function of possible fragment completions.

Conclusion

The present experiments examined aging and the effects of interpolated testing on episodic memory updating of stimuli with shared and changed features. Interpolated testing produced comparable benefits for older and younger adults partly by promoting the cross-episode integration of competing responses. This was shown by increased associations in response dependence in interpolated testing conditions, which led to proactive facilitation in recall of recent information. This association was weaker for older than younger adults, suggesting that although older adults benefitted from more opportunities to integrate responses during encoding, other cognitive deficits limited their ability to do so. This study adds to the literature suggesting that age-related memory deficits created by response competition can be reduced under conditions that engender integration of related information. Future work should explore the boundary conditions of interpolated testing benefits on episodic memory updating in older adults focusing on interactions among interpolated retrieval success, feedback, and the perceptual and conceptual dimensions on which competing information can be associated.

References

Amer, T., Wynn, J. S., & Hasher, L. (2022). Cluttered memory representations shape cognition in old age. *Trends in Cognitive Sciences*, *26*(3), 255–267. https://doi.org/10.1016/j.tics.2021.12.002

Arenberg, D. (1973). Cognition and aging: Verbal learning, memory, problem solving, and aging. In C. Eisdorfer & M. P. Lawton (Eds.), *The psychology of adult development and aging* (pp. 74–97). American Psychological Association. https://doi.org/10.1037/10044-005

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

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bartsch, L. M., & Oberauer, K. (2023). The contribution of episodic long-term memory to working memory for bindings. *Cognition*, *231*, Article 105330. https://doi.org/10.1016/j.cognition.2022.105330

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). 1–48. https://doi.org/10.18637/jss.v067.i01

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, *68*(1), 39–53. https://doi.org/10.1016/j.jml.2012.07.006

Bellezza, F. S., & Schirmann, N. J. (1975). Response dependence in simultaneously learned A-B, A-C lists. *Journal of Verbal Learning and Verbal Behavior*, *14*(1), 89–94. https://doi.org/10.1016/S0022-5371(75)80009-0

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*(4), 558–561. https://doi.org/10.1093/geronb/gbs074

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *60*(5), P223–P233. https://doi.org/10.1093/geronb/60.5.P223

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Campbell, K. L., Lustig, C., & Hasher, L. (2020). Aging and inhibition: Introduction to the special issue. *Psychology and Aging*, *35*(5), 605–613. https://doi.org/10.1037/pag0000564

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146. https://doi.org/10.1037/bul0000166

Craik, F. I. M. (2017). *Memory, attention, and aging: Selected works of Fergus I.M. Craik* (1st ed.). Routledge, Taylor & Francis Group.

Davis, S. D., & Chan, J. C. K. (2015). Studying on borrowed time: How does testing impair new learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1741–1754. https://doi.org/10.1037/xlm0000126

Delhaye, E., Folville, A., & Bastin, C. (2019). How to induce an age-related benefit of semantic relatedness in associative memory: It's all in the design. *Psychology and Aging*, *34*(4), 572–586. https://doi.org/10.1037/pag0000360

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*(7), 1049–1062. https://doi.org/10.3758/s13421-014-0425-y

Ebert, P. L., & Anderson, N. D. (2009). Proactive and retroactive interference in young adults, healthy older adults, and older adults with amnestic mild cognitive impairment. *Journal of the International Neuropsychological Society*, *15*(1), 83–93. https://doi.org/10.1017/S1355617708090115

Finn, B., & Roediger, H. L., III. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1665–1681. https://doi.org/10.1037/a0032377

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freund, J. S., & Witte, K. L. (1976). Paired-associate transfer: Age of subjects, anticipation interval, association value, and paradigm. *The American Journal of Psychology*, *89*(4), Article 695. https://doi.org/10.2307/1421467

Garlitch, S. M., & Wahlheim, C. N. (2020). The role of reminding in retroactive effects of memory for older and younger adults. *Psychology and Aging*, *35*(5), 697–709. https://doi.org/10.1037/pag0000427

Garlitch, S. M., & Wahlheim, C. N. (2021). Directing attention to event changes improves memory updating for older adults. *Psychology and Aging*, *36*(4), 475–490. https://doi.org/10.1037/pag0000503

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, *92*(10). 1–29. https://doi.org/10.18637/jss.v092.i10

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 22, pp. 193–225). Elsevier. https://doi.org/10.1016/S0079-7421(08)60041-9

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging*, *14*(1), 122–134. https://doi.org/10.1037/0882-7974.14.1.122

Hetherington, R. (1954). The Snellen chart as a test of visual acuity. *Psychologische Forschung*, *24*(4), 349–357. https://doi.org/10.1007/BF00422033

Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*(1), 70–74. https://doi.org/10.1037/h0034761

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language*, *35*(1), 32–52. https://doi.org/10.1006/jmla.1996.0002

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 3–22. https://doi.org/10.1037/0278-7393.25.1.3

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*(3), 686–700. https://doi.org/10.1037/0278-7393.27.3.686

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1282–1297. https://doi.org/10.1037/xlm0000123

Kane, M. J., & Hasher, L. (1995). Interference. In G. L. Maddox (Ed.), *The encyclopedia of aging* (2nd ed., pp. 514–516). SpringerVerlag.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. *The American Journal of Psychiatry*, *140*(6), 734–739. https://doi.org/10.1176/ajp.140.6.734

Kausler, D. H. (1994). *Learning and memory in normal aging*. Academic Press.

Kemp, P. L., Wahlheim, C. N., & Loaiza, V. M. (2023). *Testing enhances episodic memory updating in younger and older adults* [Data set]. https://doi.org/10.17605/OSF.IO/BXHWA

Kirchhoff, B. A., Anderson, B. A., Barch, D. M., & Jacoby, L. L. (2012). Cognitive and neural effects of semantic encoding strategy training in older adults. *Cerebral Cortex*, *22*(4), 788–799. https://doi.org/10.1093/cercor/bhr129

Lenth, R. (2021). *emmeans: Estimated marginal means, aka least-squares means (1.6.3)*. https://cran.r-project.org/web/packages/emmeans/index.html

Loaiza, V. M., & Srokova, S. (2020). Semantic relatedness corrects the age-related binding deficit in working memory and episodic memory. *The Journals of Gerontology: Series B*, *75*(9), 1841–1849. https://doi.org/10.1093/geronb/gbz055

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a ‘new view. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 145–162). American Psychological Association. https://doi.org/10.1037/11587-008

Malis, G. (1970). *The priority effect in the A-B, A-C paradigm* [Unpublished Doctoral Dissertation]. University of Toronto.

McKinley, G. L., & Benjamin, A. S. (2020). The role of retrieval during study: Evidence of reminding from overt rehearsal. *Journal of Memory and Language*, *114*, Article 104128. https://doi.org/10.1016/j.jml.2020.104128

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*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms* [Database]. http://w3.usf.edu/FreeAssociation/

Oberauer, K. (2019). Working memory capacity limits memory for bindings. *Journal of Cognition*, *2*(1), Article 40. https://doi.org/10.5334/joc.86

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

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*(1), 82–90. https://doi.org/10.1093/geronb/gbw066

Pastötter, B., & Bäuml, K. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging*, *34*(2), 242–250. https://doi.org/10.1037/pag0000307

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, Article 286. https://doi.org/10.3389/fpsyg.2014.00286

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*(2), 287–297. https://doi.org/10.1037/a0021801

Postman, L., & Gray, W. (1977). Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(3), 255–263. https://doi.org/10.1037/0278-7393.3.3.255

Psychology Software Tools. (2012). *[E-Prime 2.0]*. https://www.pstnet.com

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: A meta-analysis. *Psychonomic Bulletin & Review*, *26*(5), 1529–1547. https://doi.org/10.3758/s13423-019-01649-y

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Salthouse, T. A. (2011). *Theoretical perspectives on cognitive aging* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315785363

Shipley, W. C. (1986). *Shipley Institute of Living Scale* [Data set]. Western. Psychological Services. https://doi.org/10.1037/t07774-000

Stawarczyk, D., Wahlheim, C. N., Etzel, J. A., Snyder, A. Z., & Zacks, J. M. (2020). Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29346–29353. https://doi.org/10.1073/pnas.1918063117

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*(6), 1392–1399. https://doi.org/10.1037/a0013082

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, *143*(4), 1526–1540. https://doi.org/10.1037/a0036036

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, *13*(2), 181–193. https://doi.org/10.1016/S0022-5371(74)80043-5

Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition*, *42*(6), 950–964. https://doi.org/10.3758/s13421-014-0411-4

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *43*(1), 27–38. https://doi.org/10.3758/s13421-014-0455-5

Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science*, *31*(10), 1325–1339. https://doi.org/10.1177/0956797620952797

Wahlheim, C. N., & Garlitch, S. M. (2020). Adult age differences in the use of temporal and semantic context in dual-list free recall. *Psychology and Aging*, *35*(1), 143–157. https://doi.org/10.1037/pag0000425

Wahlheim, C. N., Garlitch, S. M., & Kemp, P. L. (2021). Context differentiation and remindings in episodic memory updating. *Psychology of Learning and Motivation*, *75*, 245–277. https://doi.org/10.1016/bs.plm.2021.06.001

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, *41*(1), 1–15. https://doi.org/10.3758/s13421-012-0246-9

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General*, *148*(1), 30–50. https://doi.org/10.1037/xge0000458

Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology*, *6*(2), 82–85. https://doi.org/10.1016/0361-476X(81)90035-7

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, *18*(3), 518–523. https://doi.org/10.3758/s13423-011-0085-x

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2012). A default bayesian hypothesis test for ANOVA designs. *The American Statistician*, *66*(2), 104–111. https://doi.org/10.1080/00031305.2012.695956

Winocur, G., & Moscovitch, M. (1983). Paired-associate learning in institutionalized and noninstitutionalized old people: An analysis of interference and context effects. *Journal of Gerontology*, *38*(4), 455–464. https://doi.org/10.1093/geronj/38.4.455

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*, *3*(1), Article 8. https://doi.org/10.1038/s41539-018-0024-y

Received April 26, 2022
Revision received July 22, 2023
Accepted July 30, 2023 ■