The Pretesting Effect on Older Adults' Learning of Novel Information

Authors: Wang Tangsheng, Yang Chunliang, Zhong Nian, Yang Chunliang

Date: 2020-07-04T00:00:00+00:00

Abstract

To examine whether pre-testing can effectively enhance older adults' ability to learn novel information, elderly participants were randomly assigned to an experimental group with pre-testing or a control group without pre-testing. In Experiment 1, they learned five lists of words; in Experiment 2, they memorized five lists of daily necessities; and in Experiment 3, they studied four lecture videos. The results demonstrated that older adults in the experimental group exhibited superior memory performance for novel information compared to those in the control group. It was concluded that pre-testing can significantly improve older adults' capacity to learn new information, and the 'addition-subtraction' theory, which combines interference reduction with learning engagement, provides a better explanation for the pre-testing effect on memory enhancement in older adults.

Full Text

Preamble

The Role of Forward Testing Effect in Older Adults'Learning of New Information

Tangsheng Wang a,c; Chunliang Yang b,2; Nian Zhong c
a Wuhan University of Technology, Wuhan, 430070
b Beijing Normal University, Beijing, 100875
c Wuhan University, Wuhan, 430072

Abstract

To examine whether pretesting can effectively enhance older adults' ability to learn new information, elderly participants were randomly assigned to either an experimental group with pretesting or a control group without pretesting. In Experiment 1, they learned five lists of words; in Experiment 2, they memorized five lists of daily necessities; and in Experiment 3, they studied four lecture video segments. Results showed that older adults in the experimental group consistently outperformed those in the control group on memory measures for new information. The findings suggest that pretesting significantly enhances older adults' ability to learn new information, and that a combined "addition-subtraction" theory of interference reduction and learning engagement better explains the forward testing effect in older adults' memory.

Keywords: pretesting; forward testing effect; older adults; learning engagement; interference reduction

Many countries are experiencing accelerated population aging, with individuals aged 60 and above comprising an increasingly large proportion of society. The work and life conditions of this age group have become critically important in these nations, and research, assessment, and improvement of their cognitive abilities bear significant implications for personal life, social systems, and national policy decisions.

Within the study of cognitive aging in older adults, memory decline may be among the most crucial areas. Most findings have been negative, such as limited working memory resources (Whitebourne & Whitebourne, 2014), increased risk of dementia (Hafkemeijer, Van & Rombouts, 2012), and age-related declines in episodic memory (Old & Naveh-Benjamin, 2008), autobiographical memory (Piolino, Desgranges, Benali, & Eustache, 2002), and prospective memory (Farrimond, Knight, & Titov, 2003).

Despite these negative effects of aging on memory, some positive findings have emerged. Older adults show minimal decline in semantic memory (Wiggs, Weisberg, & Martin, 2006) and self-experience memory (Gluck & Bluck, 2007). Moreover, not everyone is affected by aging in the same way (Whitebourne & Whitebourne, 2014). Encouragingly, memory training has been shown to improve cognitive function in older adults (Craik & Rose, 2012).

The present study primarily investigates whether the forward testing effect can improve memory performance in older adults, thereby laying a foundation for further enhancing their learning and memory abilities.

Numerous studies have demonstrated that testing previously learned information facilitates learning and retention of new information compared to restudying or doing nothing—this phenomenon is known as the forward testing effect (Pastötter & Bäuml, 2014; Yang, Potts, & Shanks, 2018). As Yang et al. (2018) summarized, the forward testing effect is a robust and reliable phenomenon across various materials, including vocabulary (Szpunar, McDermott, & Roediger, 2008), pictures (Pastötter, Weber, & Bäuml, 2013), foreign language translation pairs (Yang, Potts, & Shanks, 2017), face-name pairs (Weinstein, McDermott, & Szpunar, 2011), text passages (Wissman, Rawson, & Pyc, 2011), lecture videos (Szpunar, Khan, & Schacter, 2013), and drawings (Yang & Shanks, 2018). The forward testing effect is not limited to teacher-guided learning but also extends to self-regulated learning (Yang, Chew, Sun, & Shanks, 2019).

Szpunar et al. (2008) conducted a classic study demonstrating the forward testing effect with simple materials. They instructed two groups of participants (college students) to study five lists of words, each containing 18 items, presented one word at a time across lists. The experimental group took an interim test after learning each list, while the control group took an interim test only after the fifth list, engaging in no testing for lists 1-4. Results showed that the experimental group recalled twice as many words as the control group on the interim test for list 5, demonstrating the forward testing effect for simple materials. Additionally, due to proactive interference (PI), the control group made more than ten times as many errors as the experimental group on the list 5 interim test (when asked to recall list 5 words, they incorrectly recalled words from lists 1-4). Recent literature has continued to document the forward testing effect for simple materials (e.g., Nunes & Weinstein, 2012; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Weinstein, Gilmore, Szpunar, & McDermott, 2014; Yang et al., 2017).

Szpunar et al. (2008) proposed the Release from PI theory to explain the forward testing effect for simple materials. According to this theory, the list-by-list recall strategy changes the retrieval context for target words, facilitating segregation of different word lists and reducing proactive interference. Because the experimental group took interim tests, this episodic memory encoding resulted in fewer contextual competitors for new information, reducing proactive interference and thereby facilitating recall of target items (e.g., words in list 5).

In these simple learning experiments, the interim tests for lists 1-4 are called pretests (interpreted testing). They not only facilitate learning of new simple materials but also help with complex materials such as lecture videos (Jing, Szpunar, & Schacter, 2016) and article passages (Wissman et al., 2011). Szpunar et al. (2013) instructed three groups of participants to study four lecture video segments sequentially. The experimental group took an interim test after each segment, while the other two groups took an interim test only after the final segment—after studying the first three segments, they either restudied (restudy group) or solved math problems (math group). All participants could take notes and reported whether their minds were on-task while watching. Results showed the experimental group exhibited superior memory for post-test video content compared to the math and restudy groups, demonstrating the forward testing effect for complex materials. Szpunar et al. (2013) also observed that the experimental group took more notes and reported less mind wandering (MW). The forward testing effect for complex materials has been replicated in numerous recent studies (Schacter & Szpunar, 2015; Szpunar, Jing, & Schacter, 2014).

For simple materials, the Release from PI theory largely explains the forward testing effect. However, this theory struggles to account for the effect with complex materials, as proactive interference is limited in such contexts and the forward testing effect should not occur (Wissman et al., 2011). Szpunar et al. (2013) suggested that for complex materials, the forward testing effect occurs because pretesting reduces task-unrelated mind wandering. However, Yang et al. (2017) observed that participants' study motivation (time spent studying) declined continuously during learning without interim tests, but remained stable when interpolated tests were inserted. Pastötter et al. (2011) found that learners' attention to memory tasks decreased during study but was restored during interim testing. These findings support the learning engagement (LE) theory, which posits that for complex materials, the forward testing effect occurs because pretesting reverses the decline in learning engagement during cyclical study episodes, thereby facilitating learning of new complex information.

In summary, robust and reliable forward testing effects have been demonstrated across various materials. For simple materials, the effect may occur because pretesting reduces proactive interference; for complex materials, it may occur because pretesting enhances learning engagement. Importantly, these two mechanisms (interference reduction and engagement enhancement) are not mutually exclusive, and both may contribute to the forward testing effect for both simple and complex materials. Additionally, pretesting may also suppress mind wandering during learning.

Can the forward testing effect be generalized to older adults? Although memory enhancement mechanisms may differ between older and younger adults, there are compelling reasons to predict that pretesting could serve as an effective compensatory strategy to help older adults learn and remember new information. This study hypothesizes that in three upcoming experiments, pretesting will significantly improve older adults' memory performance for new materials, regardless of whether they are simple or complex (Hypothesis 1). However, the underlying mechanism is not limited to proactive interference suppression (Hypothesis 2).

Several related studies support this hypothesis, providing stable and reliable evidence for the forward testing effect. Pastötter et al. (2013) demonstrated the effect in patients with traumatic brain injury, and Pastötter and Bäuml (2019) extended it to middle-aged and older adults learning new simple materials, exploring how pretesting promotes learning of new simple items (4- to 8-letter German nouns) by reducing proactive interference. However, they did not examine motivational engagement (LE) or mind wandering (MW) mechanisms involved in complex material learning. Contrary to our hypothesis, one might speculate that the forward testing effect cannot be generalized to adults over 60, for either simple or complex materials. Regarding simple materials, Aslan and Bäuml's (2015) study of children learning simple materials seems to support this alternative speculation. They found that while interim testing essentially prevented PI in young adults and older children, it failed to reduce PI in younger children. Aslan and Bäuml (2015) suggested that younger children lack the forward testing effect for simple materials because they have difficulty controlling PI. Similar to younger children, older adults experience greater difficulty inhibiting PI than young adults (Ikier & Hasher, 2006; Ikier, Yang, & Hasher, 2008). Therefore, one might infer that the forward testing effect observed in older children and young adults cannot be directly generalized to older adults.

If older adults do show a forward testing effect for complex materials, the interference reduction theory would face challenges because older adults exhibit less mind wandering during tasks like reading passages (Krawietz, Tamplin, & Radcansky, 2012), yet they can maintain sustained attention on relevant tasks (SART; Jackson & Balota, 2012), which challenges the learning engagement theory explanation. Of course, evidence for these contrary hypotheses is relatively limited and largely indirect.

2.1 Experiment 1

2.1.1 Participants

This experiment was conducted online. Participants were recruited through Prolific Academic3 with the following restrictions: native English speakers, over 60 years old, British nationality, and current residence in the UK. Among 44,327 individuals on Prolific Academic, 127 met these criteria.

Sample size was determined based on the effect size from Yang et al. (2017, Experiment 4; Cohen’ s d = 1.12), requiring 14 participants per group to observe a significant forward testing effect ($\alpha = 0.05$; power = 0.80). To balance experimental materials, the final sample was set at 15 participants per group. Thirty older adults (17 female, 13 male) were recruited and randomly assigned to two groups: a testing group and a math group (control group), with 15 participants each. Three participants reported a history of mental illness (clinically diagnosed): one with depression in the testing group, and one with depression and one with comorbid depression and anxiety in the math group. Excluding these three participants’ data did not change the overall pattern of results, so their data were retained. Additional demographic data (age, education, and health status) were collected. At the beginning of the experiment, participants were explicitly instructed not to take notes to aid their memory. At the end, all participants were asked whether they had taken notes; all reported that they had not.

2.1.2 Materials, Design, and Procedure

The primary stimuli consisted of five lists of 18-word strings, adapted from Yang et al. (2017, Experiment 4). The order of word lists presented to participants was determined using a Latin square design.

Participants were informed that they would study five word lists sequentially and should try to remember as many words as possible, as there would be a final test covering all lists. They were also told that after each list, they would complete one minute of math problems, after which the computer would randomly decide whether to have them recall (i.e., test) the just-studied list or spend another minute on math problems. In reality, the interim tests were predetermined. The testing group took an interim test after each list (interim test recall, i.e., immediate testing for that list), while the math group solved math problems after studying the first four lists and took an interim test only after the fifth list. The experimental flowchart is shown in Figure 1 [Figure 1: see original paper].

During presentation of each list, words appeared one at a time in the center of the screen for 2 seconds each, with a 500-millisecond interval between words. After each list, both groups completed one minute of math problems (e.g., 45+62=?). During interim tests, participants had unlimited time to recall words from the current list. After all five lists were studied, both groups took a final test (final test recall) in which they were asked to recall as many words as possible from all five lists. All tests were untimed and provided no feedback on correctness.

Figure 1. Experimental procedure for the forward testing effect

The experimental group (testing group) took an immediate test after studying each block, while the control group (math group or restudy group) engaged in math problems or restudied after studying the initial blocks. Both groups took an immediate test after the final block, followed by a final comprehensive test.

Table 1 shows between-group differences on control variables for Experiment 1

Group and between-group differences	Age (years)	Education (years)	Health (5-point scale)
Testing group: Mean (SD)	66.73 (5.73)	15.13 (3.27)	3.53 (0.74)
Math group: Mean (SD)	66.13 (3.92)	15.92 (4.98)	3.33 (0.98)
Difference: t-value (p-value)	0.33 (.74)	0.63 (.53)	0.30 (0.61)

Note: Means (standard deviations) represent demographic variables for both groups, with t-values and p-values indicating the significance of differences. The same applies to all subsequent tables.

2.1.3 Results

Demographic information is presented in Table 1. No significant differences were found between groups in age, education, or health status.

Figure 2 [Figure 2: see original paper]A shows recall performance on interim tests across the five lists for both groups. A repeated-measures ANOVA within the testing group revealed that correct recall on interim tests decreased significantly across lists, $F(1, 14) = 3.80$, $p = .07$, $p^2 = 0.21$, indicating that interpolated testing did not completely prevent the decline in recall of new information as the number of studied lists increased in older adults. Most critically, the difference between groups on the list 5 immediate test was substantial. The testing group ($M = 5.87$, $SD = 1.21$) recalled approximately twice as many words as the math group (M = 3.20, SD = 2.01), with a mean difference of 2.67 words, 95% confidence interval [1.02, 4.32], t(28) = 3.31, p = .003, Cohen’ s d = 1.21, demonstrating that pretesting facilitated learning of new information in older adults.

Figure 2B shows proactive interference effects on interim tests (e.g., when asked to recall a list, participants incorrectly recalled words from previous lists). A within-group ANOVA on lists 2-5 for the testing group revealed that proactive interference increased significantly across lists, F(1, 14) = 8.15, p = 0.01, p^2 = 0.37, indicating that pretesting could not completely prevent the impact of proactive interference in older adults. Notably, the math group (M = 3.40, SD = 1.59) exhibited approximately three times the proactive interference of the testing group (M = 1.20, SD = 1.21), a difference of about 2.2 intrusions, 95% confidence interval [1.14, 3.20], t(28) = 4.26, $p < 0.001$, Cohen’ s d = 1.56, indicating that pretesting substantially reduced proactive interference.

Figure 2C shows final recall performance. On the final test, the testing group (M = 19.73, SD = 7.24) significantly outperformed the math group (M = 13.00, SD = 6.95), with a mean difference of 6.73 words, 95% confidence interval [1.43, 12.04], t(28) = 2.60, p = 0.015, Cohen’ s d = 0.95.

Figure 2. A: Recall performance on interim tests across five lists. B: Proactive interference across five interim tests. C: Number of words recalled per list on final test (Experiment 1). Error bars represent $±$1 standard deviation.

2.1.4 Summary

Although interim testing could not completely prevent the decline in memory performance as the amount of new information increased, pretesting significantly enhanced older adults’ learning and memory of new information, indicating that the forward testing effect applies to older adults. Unlike previous findings with younger participants, interim testing could not entirely eliminate proactive interference in older adults, though it substantially reduced its detrimental effects.

2.2 Experiment 2

This experiment extended Experiment 1’ s design in three ways. First, while Experiment 1 showed that interpolated testing promoted older adults’ learning compared to math problems, Experiment 2 used restudying rather than math problems as the control condition after the first four lists, with both groups taking an immediate test after list 5. Second, we assessed participants’ basic cognitive abilities to ensure that the testing group’ s superior recall on the list 5 interim test was due to the four previous interim tests rather than other factors. Third, to make the memory task more ecologically valid, we initially attempted face-surname and face-city name pairings, but pilot studies showed these materials were too difficult for older adults, so we used daily necessities instead.

2.2.1 Participants

Older adult participants were recruited through community networks familiar to the experimenters, primarily from two residential communities in Wuhan, mainland China. Sample size was determined based on the effect size from Experiment 1 (Cohen’ s d = 1.21), requiring at least 12 participants per group to observe a significant forward testing effect ($\alpha = 0.05$; power = 0.80). To balance material presentation order, we set 15 participants per group. Thirty-three older adults aged 60 and above participated; one withdrew during the experiment, leaving complete data for 32 participants (24 female). They were randomly assigned to two groups: 16 in the testing group and 16 in the restudy group (control). One participant reported a history of depression (clinically diagnosed); including or excluding her data did not change the overall pattern of results. All participants were native Chinese speakers, tested individually in quiet environments, and received 50 RMB upon completion.

2.2.2 Materials, Design, and Procedure

Fifty pictures of common supermarket items (e.g., apple, toothbrush, shampoo) were selected from Baidu Wenku4 as primary stimuli. Items were randomly assigned to five groups of 10 each, with the item name displayed below each picture in 2-3 Chinese characters. A Latin square design balanced the order of materials across participants.

4 See https://wenku.baidu.com/view/9ef45d97915f804d2b16c1f0.html

At the experiment’ s outset, all participants reported their age, gender, and years of education. They were then asked whether they had ever had a mental illness and, if so, what type. They also rated their health status on a scale from 1 (very poor) to 5 (very good).

Participants were then instructed to imagine they were preparing to go grocery shopping and would be given five shopping lists to remember as best as possible. Other instructions and procedures were identical to Experiment 1, except that the control condition was restudying: after studying lists 1-4, the restudy group spent one minute on math problems and then restudied the list (whereas the testing group took an interim test after each list). Each item was presented individually for 2 seconds. After the final test, all participants completed a vocabulary ability test, a digit span task to assess working memory capacity (WMC), and a cognitive processing speed task.

The vocabulary task was drawn from the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), using 20 two-character words from its Chinese version (Gong, 1992). The experimenter read each word aloud in random order, and participants had to locate the word in the list and explain its meaning. Successfully locating the word earned 1 point, and correct explanation earned another point, for a maximum score of 40.

Working memory capacity was assessed via the digit span task (Grégoire & Van

Der Linden, 1997).

For the cognitive processing speed task, participants judged as quickly and accurately as possible whether two strings of numbers were identical (Wang, Shen, Peng, Tang, & Zhang, 2005). On each trial, two 9-digit strings appeared on screen. If all nine digits matched, participants pressed a “same” key; if any digit differed, they pressed a “different” key. There were 18 such trials.

Table 2 shows between-group differences on control variables for Experiment 2

Group and between-group differences	Age (years)	Education (years)	Vocabulary (total 40)	Processing accuracy (%)	Processing speed (5-point scale)	Health (5-point scale)
Testing group: Mean (SD)	66.13 (4.93)	9.93 (3.28)	36.87 (3.31)	92.96 (7.99)	6.54 (1.11)	3.28 (0.46)
Restudy group: Mean (SD)	66.13 (4.24)	9.25 (2.29)	35.94 (3.15)	90.98 (10.32)	6.29 (1.81)	3.11 (0.77)
Difference: t-value (p-value)	0.00 (1.00)	0.68 (0.51)	0.80 (.43)	0.63 (0.32)	0.46 (0.65)	0.30 (0.61)

Note: Means (standard deviations) represent demographic variables and basic cognitive abilities for both groups, with t-values and p-values indicating the significance of differences.

2.2.3 Results

Demographic information and basic cognitive abilities are presented in Table 2. No significant differences were found between groups in age, education, health status, vocabulary ability, working memory, processing accuracy, or processing speed.

Figure 3 [Figure 3: see original paper]A shows recall performance for the testing group. A repeated-measures ANOVA within the testing group revealed that correct recall on interim tests decreased significantly across lists, F(1, 14) = 17.87, p = .001, p^2 = 0.56. On the list 5 immediate test, the testing group (M = 5.13, SD = 1.41) recalled twice as many items as the restudy group (M = 2.38, SD = 1.63), with a mean difference of 2.76 items, 95% confidence interval [1.64, 3.88], t(29) = 5.03, $p < .001$, Cohen’ s d = 1.80.

Figure 3B shows proactive interference reduction effects on interim tests. A within-group ANOVA on lists 2-5 for the testing group revealed that proactive interference increased significantly across lists, F(1, 14) = 20.70, $p < 0.001$, p^2 = 0.60. On the list 5 interim test, the restudy group (M = 3.06, SD = 2.11) exhibited approximately three times the proactive interference of the testing group (M = 0.93, SD = 0.96), a difference of about 2.13 items, 95% confidence interval [0.91, 3.35], t(29) = 3.57, p = 0.001, Cohen’ s d = 1.30.

Figure 3C shows final recall performance. On the final test, the testing group (M = 21.80, SD = 4.86) significantly outperformed the restudy group (M = 16.13, SD = 5.26), with a mean difference of 5.68 items, 95% confidence interval [1.95, 9.45], t(29) = 3.11, p = 0.004, Cohen’ s d = 1.12.

Figure 3. A: Recall performance on interim tests across five lists. B: Proactive interference across five interim tests. C: Number of items recalled per list on final test (Experiment 2). Error bars represent $±$1 standard deviation.

2.2.4 Summary

For memory of everyday items, pretesting also significantly enhanced older adults’ learning and memory of new information, proving even more effective than restudying. This again demonstrates the important facilitative role of the forward testing effect in older adults’ learning of new information. However, as the amount of material to be remembered increased, older adults’ recall performance declined significantly across lists. Although interpolated testing partially prevented proactive interference, it could not stop the overall downward trend in memory performance across lists.

2.3 Experiment 3

Experiments 1 and 2 demonstrated that pretesting can improve older adults’ memory for new simple materials, establishing the forward testing effect for simple learning in older adults. Experiment 3 examines whether this effect extends to older adults’ learning of complex materials.

2.3.1 Participants

Participants were recruited primarily from Jianghan Oilfield University for the Elderly and from a community in Xiangtan, Hunan Province. Fifty-two older adults (aged 60 and above) volunteered. Three did not return within 24 hours for the basic cognitive ability test, leaving 49 participants (38 male) who completed the study. They were randomly assigned to two groups: 25 in the experimental group and 24 in the restudy group (control). None had participated in Experiment 2. All participants completed demographic and basic cognitive ability tests. All were native Chinese speakers, tested individually in soundproof rooms, and received 50 RMB (or equivalent) upon completion.

2.3.2 Materials, Design, and Procedure

The primary stimuli were lecture video segments about doctor-patient relationships, obtained from the NetEase Open Course website (http://open.163.com/movie/2015/2/s/8/MAH8PH40M
The lecture was divided into four segments, each 4 minutes long. For simplicity,

each segment consisted of 15 short sentences, with each sentence containing one key word as the test target.

Sample size was determined based on the effect size from Szpunar et al. (2013, Experiment 2; Cohen' s d = 1.06), requiring 24 participants per group to observe a significant forward testing effect ($\alpha = 0.05$; power = 0.95). At the experiment' s outset, participants reported demographic information. They were then instructed to study four lecture video segments, being told that there would be a final test on all four segments after the fourth segment, and that the computer would randomly decide whether to give them a test (with 30 seconds of math problems) or have them restudy after each segment. In reality, the experimental group took an interim test after each segment, while the control group restudied segments 1-3 and took an interim test only after segment 4.

During interim tests, a text summary of the just-studied segment appeared on screen, containing 15 sentences with one blank per sentence (e.g., "In medical disputes, doctors believe they should first {___} themselves"). Participants filled in the blanks with their answers. No time limit was imposed per question, and blanks could be left empty if the answer was unknown. After each interim test, participants received correct feedback, with the summary redisplayed. Correct answers were marked in red and underlined, and participants had 60 seconds to study this feedback. For segments 1-3, the restudy group restudied the video segment, with the text summary (containing the same 15 sentences) displayed for 150 seconds. The 15 target sentences were marked in red and underlined to indicate which words would be tested later. The restudy group took the same test, viewed feedback, and completed math problems for segment 4 as the experimental group.

After the segment 4 test, both groups took a final test consisting of 20 fill-in-the-blank questions (5 from each of the 4 segments), all presented on one screen. Only 20 questions were used rather than all 60 because pilot participants requested shorter experiments. Participants generally returned within 24 hours to complete the cognitive ability tests, identical to those in Experiment 2.

2.3.3 Results

Demographic information and basic cognitive abilities are presented in Table 3 . No significant differences were found between groups in age, education, health status, vocabulary ability, working memory, processing accuracy, or processing speed.

Two experiment assistants blind to the study' s purpose scored participants' memory tests. One assistant scored all memory tests, while the other scored only the interim test for segment 4. Their ratings were highly consistent: 95.9% agreement on scoring of segment 4 interim tests. Discrepant items were rescored by the experimenter.

For segments 1-3, the testing group' s recall rates were 9.00 (SD = 2.42), 9.08

(SD = 3.09), and 7.76 (SD = 2.52), respectively. The two groups' performance on the segment 4 interim test was critical (see Figure 4 [Figure 4: see original paper]A). On the segment 4 interim test, the testing group (M = 6.32; SD = 3.11) recalled twice as many items as the restudy group (M = 3.50; SD = 2.17), with a mean difference of 2.82 items, 95% confidence interval [1.28, 4.36], t(47) = 3.68, p = 0.001, Cohen' s d = 1.05, demonstrating that pretesting significantly improved older adults' learning of new complex materials. As shown in Figure 4B, on the final comprehensive test, the testing group (M = 15.84; SD = 2.29) also outperformed the restudy group (M = 12.21; SD = 3.87), with a difference of 3.63 items, 95% confidence interval [1.81, 5.45], t(47) = 4.02, $p < 0.001$, Cohen' s d = 1.14.

Table 3 shows between-group differences on control variables for Experiment 3

Group and between-group differences	Age (years)	Education (years)	Vocabulary (total 40)	Processing accuracy (%)	Processing speed (5-point scale)	Health (5-point scale)
Testing group: Mean (SD)	66.44 (6.44)	9.08 (3.11)	37.08 (3.21)	96.22 (5.25)	6.37 (2.13)	3.20 (0.64)
Restudy group: Mean (SD)	66.21 (6.20)	9.21 (3.11)	36.21 (4.23)	96.30 (4.54)	5.73 (1.22)	3.17 (0.83)
Difference: t-value (p-value)	0.14 (.89)	-0.14 (.89)	0.81 (0.42)	-0.05 (0.96)	0.48 (0.64)	0.48 (0.64)

Note: Means (standard deviations) represent demographic variables and basic cognitive abilities for both groups, with t-values and p-values indicating the significance of differences.

Figure 4. A: Interim test performance for segment 4 video. B: Final test performance (Experiment 3). Error bars represent $±$1 standard deviation.

2.3.4 Summary

These results demonstrate that the forward testing effect for complex materials also extends to older adults. Because complex material learning does not involve proactive interference to the same extent as simple material learning, Experiment 3 only compared interim and final test performance. Interestingly, the testing group' s interim test performance for segment 2 was not worse than for segment 1, but after segment 2, recall accuracy showed a clear declining trend, consistent with the first two experiments. This suggests that memory load in older adults should be given full consideration in experimental design and requires corresponding theoretical explanation.

3 General Discussion and Conclusion

In summary, the forward testing effect for learning new information extends to older adults for both simple and complex materials. These results likely reflect different psychological mechanisms triggered by pretesting. For simple materials, pretesting may primarily prevent the buildup of proactive interference (PI), while for complex materials, it may primarily reverse the decline in learning engagement during the study process. Together, these mechanisms produce the forward testing effect in older adults' learning of new information.

Specifically, Experiment 1 demonstrated that interim testing during simple material learning effectively reduced proactive interference from previously studied items—decreasing proactive interference by approximately threefold. To demonstrate that this interference reduction was not merely due to repeated study opportunities provided by testing, Experiment 2' s control group restudied after each list, yet the testing group still showed superior memory for new items, maintaining a threefold reduction in proactive interference relative to the control group.

Although the forward testing effect has been widely documented, previous research has primarily involved college students (Yang et al., 2018; Chan, Meissner, & Davis, 2018). At the time this study was conducted, no published reports had examined this specific age group (over 60). The primary aim was to verify whether pretesting could serve as a compensatory strategy to improve learning of new information in older adults (aged 60+). As discussed earlier, the forward testing effect for simple and complex materials may be driven by two mechanisms: interference reduction and learning engagement. Experiments 1 and 2 examined the interference reduction mechanism for simple materials, finding that this theory only partially explains the effect in older adults: even with pretesting, older adults showed a significant declining trend in memory performance across lists. In contrast, Szpunar et al. (2008, Experiment 4) found that with interpolated testing, earlier materials had no impact on later recall—recall remained consistent across lists and proactive interference was essentially eliminated. This suggests that for older adults in Experiments 1 and 2, the gradual decline in learning and memory performance cannot be attributed solely to proactive interference. Therefore, we speculate that interference reduction theory cannot fully explain the facilitative effect of pretesting in older adults. To test this, Experiment 3 extended the forward testing effect to complex materials (lecture videos), a context where proactive interference is generally considered limited and only differences between testing and restudy groups should exist. Yet the testing group' s recall was nearly double that of the restudy group, further confirming that additional theoretical mechanisms beyond interference reduction are needed to explain the forward testing effect in older adults.

The inability of interference reduction theory to fully explain the forward testing effect across our three experiments likely relates to participants' age. Aslan and Bäuml (2015) found that the forward testing effect does not extend to younger children because, for them, interim testing cannot reduce proactive interference regardless of its presence. Although older adults differ from younger children—the forward testing effect does facilitate their learning—even with interpolated testing, the declining trend in recall across lists cannot be prevented. Clearly, interference reduction theory cannot explain the forward testing effect across all age groups, particularly younger children and older adults.

Previous research generally holds that proactive interference does not occur in complex material learning (Wissman et al., 2011), so an alternative explanation for the forward testing effect with complex materials is that pretesting reduces task-unrelated mind wandering (MW; Szpunar et al., 2013). Some studies suggest that older adults exhibit less MW than younger adults (Krawietz et al., 2012), with age positively correlated with reduced MW (Seli, Maillet, Smilek, Oakman, & Schacter, 2017). Our finding that pretesting significantly improves older adults' learning of complex materials makes it difficult to attribute the forward testing effect solely to reduced MW.

Furthermore, although some studies find that older adults report fewer MW episodes than younger adults, certain behavioral indices of MW are higher in older adults (Zavagnin, Borella, & De Beni, 2014). According to Jordano and Touron (2017), older adults are more susceptible to task-related MW than younger adults, and they found that older adults are more influenced by motivation, with motivational intensity negatively correlated with unintentional MW. Thus, the facilitative effect of pretesting likely relates closely to older adults' motivation for learning complex materials—stronger learning motivation leads to less unintentional MW and more task-related MW. Pretesting in our study may have suppressed unintentional MW, but more likely suppressed task-related MW. According to Yang et al. (2019), pretesting during learning can improve participants' learning engagement, strengthening their involvement. Our older adults' memory data across three experiments strongly support the learning engagement (LE) theory' s explanation of the forward testing effect.

Learning engagement theory can explain many phenomena related to the forward testing effect. First, pretesting leads people to spend more time encoding new information than control groups (Yang et al., 2017). Second, Pastötter et al. (2011) found that for control groups, alpha wave activity increased significantly across several study blocks, whereas no such increase occurred for testing groups, suggesting that interim testing changed the context of studied materials (target search scope), resetting memory encoding for subsequent materials to be as efficient as for earlier blocks. Third, interim testing facilitates learning and improves memory even when material types change between blocks (Yang et al., 2019), making motivational explanations more compelling.

Of course, the learning engagement theory' s explanation of the forward testing effect' s psychological mechanisms is still developing. Weinstein et al. (2014) proposed that interim testing creates expectations about the experimental procedure, thereby enhancing motivation to encode and retrieve new information. Cho, Neely, Croco, and Vitrano (2017) suggested that encoding and retrieval failures during pretesting create dissatisfaction, motivating participants to expend more effort on encoding and retrieval in subsequent tests.

In conclusion, we propose that only by combining interference reduction theory and learning engagement theory, while considering characteristics of participants' memory abilities, can the forward testing effect in older adults' learning and memory be effectively explained. We term this the "addition-subtraction" theory of the forward testing effect.

First, integrating interference reduction theory with the memory characteristics of older adults described in the introduction, the testing group' s superior memory for new materials in Experiment 1 may be because pretesting' s episodic memory encoding reduced contextual competitors for new information, so reduced proactive interference partially facilitated the forward testing effect. However, because older adults' episodic memory generally declines, even interpolated testing cannot completely prevent the overall declining trend in recall across lists. That is, older adults' general "subtraction" of episodic memory ability and the "subtraction" of interference competitors jointly influence final memory performance in Experiment 1.

Second, although older adults' prospective memory declines, pretesting may simplify the context for encoding and retrieving target materials, reducing prospective memory load and enhancing their confidence and motivation for learning engagement, ultimately improving their everyday prospective memory (Experiment 2). However, because older adults' prospective memory ability generally declines, even interpolated testing cannot completely prevent the overall declining trend in memory for new information. That is, older adults' general "subtraction" of prospective memory ability, the "subtraction" of contextual competitors, and the "addition" of learning engagement motivation jointly influence final memory performance in Experiment 2.

Finally, only by combining interference reduction theory and learning engagement theory can the forward testing effect for complex materials in older adults be fully explained (Experiment 3). According to interference reduction theory, the forward testing effect for complex materials occurs because of reduced mind wandering. However, as noted earlier, older adults exhibit less task-unrelated MW but more task-related MW, creating difficulties for interference reduction theory' s explanation. Nevertheless, according to Jordano and Touron (2017), task-related MW in older adults is triggered by age-related stereotypes, with these negative aging stereotypes focusing attention on self-evaluative threats in the learning environment. In Experiment 3, the "random" nature of interim testing reduced older adults' self-evaluative threat, thereby reducing task-related MW, while pretesting simultaneously enhanced their learning engagement, resulting in better test performance than would occur with task-related MW alone. Thus, pretesting both enhanced learning engagement and suppressed interfering mind wandering, jointly improving memory for new complex materials.

Based on our "addition-subtraction" theory, we can speculate that cultural factors and activation of aging stereotypes would influence the effectiveness of interim testing for older adults' learning of new information. Levy and Langer (1994) found that negative aging stereotypes affect older adults' memory performance. Because Chinese culture contains fewer negative aging stereotypes, memory performance is better than among American older adults. Although our study included older adults from different cultures, we did not deeply examine how cultural and self-evaluation factors (aging stereotypes) influence the forward testing effect, which should be explored in future research.

This study' s advancement in understanding the forward testing effect in older adults' learning and memory holds promise for ameliorating age-related memory decline. Kielar et al. (2019) found that patients with primary progressive aphasia lose speaking or language comprehension abilities due to functional deficits in certain brain regions that show no structural damage on MRI, suggesting that functional rather than structural deficits underlie dementia symptoms. That is, dementia may stem from functional impairments in daily life rather than structural brain pathology, suggesting a new solution: using periodic retrieval practice as a memory training strategy for older adults to compensate for cognitive disadvantages, potentially delaying or even preventing dementia onset and progression.

We thank all the older adults and research assistants who participated in this study, including but not limited to Yan Wang, Fan Yang, Fangqi Cheng, Yujin Liu, and Chongyang Liu. We also thank the reviewers for their detailed comments and suggestions, which greatly improved the quality of this paper.

6 References

Aslan, A., & Bäuml, K. H. T. (2015). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science, 19*(6), 992-998.

Chan, J., Meissner, C., & Davis, S. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin, 144*(11), 1111-1146.

Cho, K. W., Neely, J. H., Croco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology, 70*(7), 1211-1235.

Craik, F. I. M., & Rose, N. S. (2012). Memory encoding and aging: A neurocognitive perspective. *Neuroscience and Biobehavioral Reviews, 36*(7), 1729-1739.

Farrimond, S. J., Knight, R. G., & Titov, N. (2003). Prospective remembering and aging. *Archives of Clinical Neuropsychology, 18*(7), 769-769.

Gluck, J., & Bluck, S. (2007). Looking back across the life span: A life story account of the reminiscence bump. *Memory and Cognition, 35*(8), 1928-1939.

Gong, Y. (1992). *Revised Wechsler adult intelligence scale manual in China.* Hunan Map Press. [龚耀先 (1992). 中国修订韦氏成人智力量表手册. 湖南地图出版社]

Grégoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition, 4*(2), 140-149.

Hafkemeijer, A., Van, D. G. J., & Rombouts, S. A. (2012). Imaging the default mode network in aging and dementia. *Biochimica et Biophysica Acta, 1822*(3), 431-441.

Ikier, S., & Hasher, L. (2006). Age differences in implicit interference. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 61*(5), 278-284.

Ikier, S., Yang, L., & Hasher, L. (2008). Implicit proactive interference, age, and automatic versus controlled retrieval strategies. *Psychological Science, 19*(5), 456-461.

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging, 27*(1), 106-116.

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied, 22*(3), 305-318.

Jordano, M. L., & Touron, D. R. (2017). Stereotype threat as a trigger of mind-wandering in older adults. *Psychology and Aging, 32*(3), 307-313.

Kielar, A., Shah-Basak, P. P., Deschamps, T., Jokel, R., & Meltzer, J. (2019). Slowing is slowing: Delayed neural responses to words are linked to abnormally slow resting state activity in primary progressive aphasia. *Neuropsychologia, 126*(6), 331-347.

Krawietz, S. A., Tamplin, A. K., & Radvansky, G. A. (2012). Aging and mind wandering during text comprehension. *Psychology and Aging, 27*(4), 951-958.

Levy, B., & Langer, E. (1994). Aging free from negative stereotypes: Successful memory in China among the American deaf. *Journal of Personality and Social Psychology, 66*(6), 989-997.

Nunes, L. D., & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory, 20*(2), 138-154.

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.

Pastötter, B., & Bäuml, K. H. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5*(4), 1-5.

Pastötter, B., & Bäuml, K.-H. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging, 34*(2), 242-250.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. H. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 287-297.

Pastötter, B., Weber, J., & Bäuml, K. H. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology, 27*(2), 280-285.

Piolino, P., Desgranges, B., Benali, K., & Eustache, F. (2002). Episodic and semantic remote autobiographical memory in ageing. *Memory, 10*(4), 239-257.

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology, 1*(1), 60-71.

Seli, P., Maillet, D., Smilek, D., Oakman, J. M., & Schacter, D. L. (2017). Cognitive aging and the distinction between intentional and unintentional mind wandering. *Psychology and Aging, 32*(4), 315-324.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition, 3*(3), 161-164.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110*(16), 6313-6317.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392-1399.

Wang, D., Shen, J., Peng, H., Tang, D., & Zhang, L. (2005). The model of educational effect on old adult’ s cognition. *Acta Psychologica Sinica, 37*(4), 511-516. [王大华, 申继亮, 彭华茂, 唐丹, & 张凌. (2005). 教育水平对老年人认知能力的影响模式. 心理学报, 37(4), 511-516]

Wechsler, D. (1955). *Manual for the Wechsler adult intelligence scale.* Oxford, England: Psychological Corp.

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 1039-1048.

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin and Review, 18*(3), 518-523.

Whitebourne, S. K., & Whitebourne, S. B. (2014). *Adult development and aging: Biopsychosocial perspectives* (5th ed.). USA: John Wiley & Sons.

Wiggs, C. L., Weisberg, J., & Martin, A. (2006). Repetition priming across the adult lifespan—The long and short of it. *Aging, Neuropsychology and Cognition, 13*(3-4), 308-325.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin and Review, 18*(6), 1140-1147.

Yang, C., Chew, S., Sun, B., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning. *Journal of Educational Psychology, 111*(5), 809-826.

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied, 23*(3), 263-277.

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*(8), 1-9.

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 44*(3), 485-492.

Zavagnin, M., Borella, E., & De Beni, R. (2014). When the mind wanders: Age-related differences between young and older adults. *Acta Psychologica, 145*, 54-64.

Note: Figure translations are in progress. See original paper for figures.