Cognitive Development 59 (2021) 101079

Contents lists available at ScienceDirect

Cognitive Development

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

Age difference in the forward testing effect: The roles of strategy change and release from proactive interference

Xixi Dang, Chunliang Yang, Yinghe Chen *

School of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, 100875, China

A R T I C L E I N F O

Keywords:
Forward testing effect
Age difference
Semantic processing
Strategy-change
Proactive interference

A B S T R A C T

Testing of previously studied information potentiates subsequent learning of new information, a phenomenon referred to as the *forward testing effect* (FTE). The current study aimed to investigate whether semantic processing strategy-change and release-from-PI (proactive interference), mediate the FTEs in adults and children. Third-grade children and adults were instructed to study four categorical word lists, then either restudied or were tested following studying each of Lists 1–3, and took an interim test on List 4. Results demonstrated that the FTE generalized to children and adults. More importantly, a moderated, multiple mediation analysis revealed that age moderated the indirect effect of interim task (test *vs.* restudy) on List 4 recall performance through semantic processing strategy-change. Specifically, the mediation effect of semantic processing strategy-change was only detected in adults but not in children. By contrast, the mediation effect of release-from-PI was evident for both children and adults. Furthermore, release-from-PI, by comparison with semantic processing strategy-change, contributed more importantly to the FTE. These findings provide novel insights into the development trajectory of mechanisms underlying the FTE.

1. Introduction

Retrieval practice of studied information cannot only enhance its long-term retention (Roediger & Karpicke, 2006; Yang, Luo, Vadillo, Yu, & Shanks, 2020), but also facilitate learning and retrieval of new information (Pastötter & Bäuml, 2014; Yang, Potts, & Shanks, 2018). These two kinds of testing benefits are referred to as the *backward testing effect* (BTE) and the *forward testing effect* (FTE), respectively. The current study focuses on the latter — the FTE. The FTE is typically investigated by using a multi-list or multi-section learning paradigm. For example, participants are asked to memorize five lists of words. For Lists 1–4, participants in the test condition take a recall test after studying each list, while those in the control condition are not tested. Then participants in both conditions study List 5 and take a test on it. The difference in List 5 recall between the two conditions represents the influence of testing on new learning (Szpunar, McDermott, & Roediger, 2008). The FTE is observed when participants in the test condition successfully recall more List 5 words than those in the no-test condition.

The FTE has been established as a robust phenomenon by using different types of study materials and in different experimental settings (Yang & Shanks, 2018; for reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018), and this effect generalizes to individuals with different levels of working memory capacity and test anxiety (Yang, Sun et al., 2020). However, two major limitations exist in this research area. First, although the FTE has been repeatedly established as a robust phenomenon and

* Corresponding author.
E-mail address: chenyinghe@bnu.edu.cn (Y. Chen).

https://doi.org/10.1016/j.cogdev.2021.101079
Received 17 November 2020; Received in revised form 14 June 2021; Accepted 16 June 2021

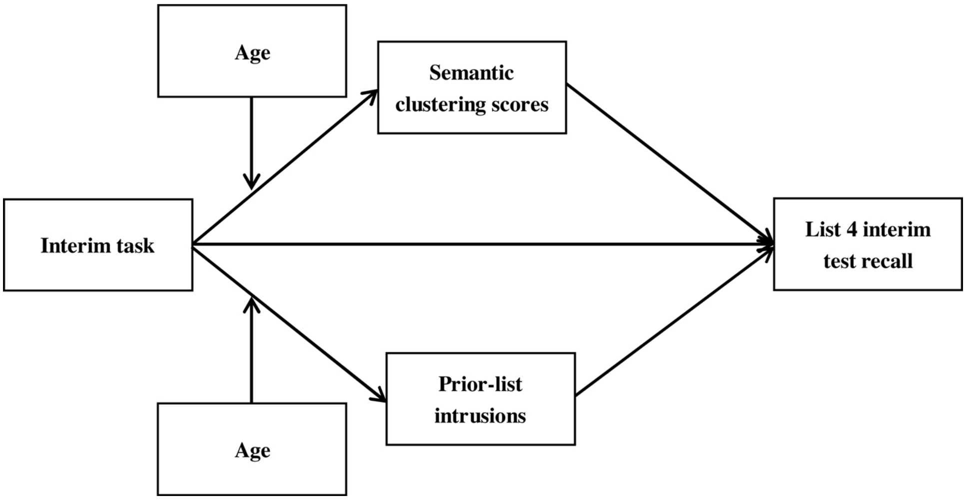

Fig. 1. Conceptual model on the relation of interim task (test *vs.* restudy) and list 4 interim test recall. The relation of interim task to list 4 interim test recall was hypothesized to be mediated by semantic clustering scores and prior-list intrusions; these mediated relations were in turn hypothesized to be moderated by age (children *vs.* adults).

of practical significance, its cognitive underpinnings remain elusive (for a detailed discussion, see Yang et al., 2018). Second, the majority of previous studies explored the FTE in adults, but little is known about whether the FTE exists in children. Thus, the main goals of the current study are to explore (1) whether the FTE generalizes to children and (2) whether different mechanisms are responsible for children's and adults' FTEs.

To our knowledge, only one study has investigated the FTE in children, which showed that the effect generalizes to older children (average age = 8.8 years) and adults but not to younger children (average age = 6.7 years) (Aslan & Bäuml, 2016). The age difference between the children's and adults' FTEs may be explained by the *release-from-proactive interference* (PI) account (Szpunar et al., 2008). According to this account, interim tests induce context changes, which in turn reduce the buildup of PI and improve retrieval of new information. Specifically, participants study some items, which are associated with a mental context of study (Context S). The mental context of these items is updated by an interim test, and hence the tested items are associated with both a study and a retrieval context (Contexts S and R) (Karpicke, Lehman, & Aue, 2014). When participants study new items and are instructed to recall these items, context differences help participants to differentiate (1) the new (only associated with Context S) and (2) previous items (associated with both Contexts S and R), which in turn reduces impairment from PI.

Subsequent studies, using either single items (Bäuml & Kliegl, 2013; Weinstein, Gilmore, Szpunar, & McDermott, 2014; Yang, Potts, & Shanks, 2017) or paired-associate items (Cho, Neely, Crocco, & Vitrano, 2017; Weinstein, McDermott, & Szpunar, 2011; Yang, Chew, Sun, & Shanks, 2019) as their principal stimuli, provide consistent evidence supporting the release-from-PI account. Thus, the release-from-PI theory may account for both children's and adults' FTEs, and the reduction of the effect in older children may derive from their deficits in inhibition of PI (Aslan & Bäuml, 2016). However, because Aslan and Bäuml did not conduct mediation analyses, it is premature to conclude that the difference between children's and adults' FTEs is driven by release-from-PI.

Another possible explanation of age difference in the FTE is the *strategy change* account (Cho et al., 2017; Yang et al., 2018). This account assumes that interim testing enhances new learning beyond context change: testing may enhance new learning by offering participants an opportunity to switch to more effective strategies in subsequent encoding and/or retrieval processes (Yang et al., 2018). Semantic processing (clustering) is one of the most important strategies for verbal recall (Craik & Tulving, 1975). After studying a list of semantically related words in a randomized order, participants tend to recall them in an organized manner by clustering related items together (Bousfield, 1953; Bousfield, Cohen, & Whitmarsh, 1958; Zaromb & Roediger, 2010). The positive effect of testing on semantic processing strategy-change has been documented in adults' FTE. For instance, Chan, Manley, Davis, and Szpunar (2018) found that interim testing induced superior semantic clustering, when compared to the no-test and restudy conditions. Moreover, semantic clustering is positively related to free recall performance (Chan, Manley, & Ahn, 2020). Thus, semantic processing strategy-change may contribute importantly to adults' FTE.

Previous studies demonstrated that testing has different influences on children's and adults' semantic processing strategies. For instance, Zaromb and Roediger (2010) observed that although testing boosts categorical clustering of word lists for adults, it fails to enhance strategic clustering in children aged 7–9 (Lipowski, Pyc, Dunlosky, & Rawson, 2014). Indeed, children aged 6–7 do not spontaneously utilize mnemonic strategies and cannot be prompted to do so (Bjorklund, Ornstein, & Haig, 1977; Wimmer & Tornquist, 1980). Although children aged 8–9 are capable of employing semantic processing strategies, they only shift to these strategies when they are directly trained (Gaskill & Murphy, 2004) or explicitly instructed (Schleepen & Jonkman, 2012; Schneider & Pressley, 1997) to do so. The above discussion suggests that testing can optimize adults' but may not be able to affect children's semantic processing strategies. Accordingly, it is reasonable to assume that semantic processing strategy-change may play an important role in adults' but not in children's FTE, and the reduced FTE for children derives from the fact that interim testing fails to optimize their semantic processing strategies.

In summary, both release-from-PI and semantic processing strategy-change appear to be likely explanations for the FTE and how it varies by age. However, it is unknown (1) whether these two mechanisms play causal roles in the FTE, (2) which mechanism contributes more importantly than the other, and (3) whether they are responsible for the FTE's age difference. Thus, the current study aims to fill these gaps.

Other theoretical accounts, including those proposing that the FTE is induced by enhanced test expectancy (*i.e.*, the test expectancy account, Weinstein et al., 2014), unsatisfactory retrieval failures (*i.e.*, the retrieval effort account, Cho et al., 2017), encoding reset (*i.e.*, the reset-of-encoding account, Pastötter, Engel, & Frings, 2018) and so on (for a review, see Yang et al., 2018), are not of main interest in the current study and hence are not discussed further.

The current study was motivated to explore the developmental trajectory of the FTE and its underlying mechanisms. Specifically, the current study targeted to examine (1) whether (and if yes, to what extent) release-from-PI and semantic processing strategy-change contribute to children's and adults' FTEs, (2) which mechanism contributes more importantly than the other, and (3) which mechanisms are responsible for the development trajectory of the FTE. Following Aslan and Bäuml (2016), we recruited third-grade children and adults, and employed four lists of categorical words as principal stimuli. In each age group, we compared prior-list intrusions and semantic clustering scores (Chan, Manley et al., 2018;Roenker, Thompson, & Brown, 1971) between a test and a restudy condition.

Based on the results reported by Aslan and Bäuml (2016), we expected to observe a robust FTE for adults, a reduced FTE for third-grade children (Hypothesis 1). More importantly, a moderated, multiple mediation model was applied to unpack three important questions: whether release-from-PI (indexed by prior-list intrusions) and semantic processing strategy (indexed by semantic clustering scores) play mediating roles in the FTE and whether their mediating effects are moderated by age. The conceptual model is depicted in Fig. 1.

As discussed above, children may be unable to shift their semantic processing strategies according to previous test experience (Gaskill & Murphy, 2004; Schleepen & Jonkman, 2012). Accordingly, it is reasonable to assume that the contribution of semantic processing strategy-change is moderated by age, such that the FTE is mediated by semantic clustering scores for adults but not for children (Hypothesis 2). Aslan and Baulm (2016) found that interim testing reduced PI for both children and adults. Therefore, it is rational to expect that release-from-PI may contribute to both adults' and children's FTEs (Hypothesis 3).

2. Method

2.1. Participants

We conducted a power analysis to estimate the required sample size using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). The effect sizes (*i.e.*, Cohen's *ds*) of the FTE, reported by Aslan and Bäuml (2016), ranged from 0.74 to 1.28. Power analysis results showed that 14–40 participants per condition were required to observe a significant FTE (two-tailed $\alpha = 0.05$) at 0.90 power. Because in the current study we also planned to apply mediation models to determine its underlying mechanisms, we hence decided to increase the sample size to 45 in each group. Correspondingly, 90 children ($M = 8.97$ years old, $SD = 0.55$; 48 female), and 90 adults ($M = 21.67$ years old, $SD = 1.87$; 45 female) were recruited. For both children and adults, 45 of them were randomly assigned into a test group, and the other 45 were divided into a restudy group.

All children were native Chinese speakers recruited from an elementary school in Shandong province of China, and adults were students recruited from Beijing Normal University. They reported normal or corrected-to-normal vision and had no neurological or psychiatric diseases (based on self-reports for adults and caregivers' reports for children). Informed consent was obtained from either participant themselves (for adults) or their caregivers (for children). The study was approved by the Ethics Committee of Beijing Normal University. Each adult participant received 15 RMB and each child participant received a set of stationery as compensation.

2.2. Materials

We conducted a pilot study to develop categorical word lists for the current study (see Appendix A. Table A1 for all stimuli). Three lists (*i.e.*, animals, fruits, and human body parts) were used in the current study. The 8 exemplars in each category were divided into four groups, with 2 words per group. These category dyads were then assembled to form four study lists, with 6 words per list. There was no significant difference in word frequency between categories ($ps > .05$).

2.3. Experiment design and procedure

The experiment involved a 2 (age: children *vs*. adults) × 2 (interim task: test *vs*. restudy) between-subjects design. All participants were tested individually in a sound-proofed room.

Participants were informed that the experiment was designed to test their memory and mathematical abilities. Specifically, they would listen to four lists of words and each list would be presented twice in a different order. They were encouraged to remember as many words as they could, in preparation for a final cumulative test in which all words from all lists would be tested. They were also informed that, after studying each list and solving math problems for 30 s, the computer program would randomly decide whether or not to give them a short test.

Following the instructions, the 6 words in each list were presented orally by the experimenter in a random order, 2 s each, and with a 2 s inter-stimulus interval. After studying each list, participants solved as many math problems as they could in 30 s. Immediately

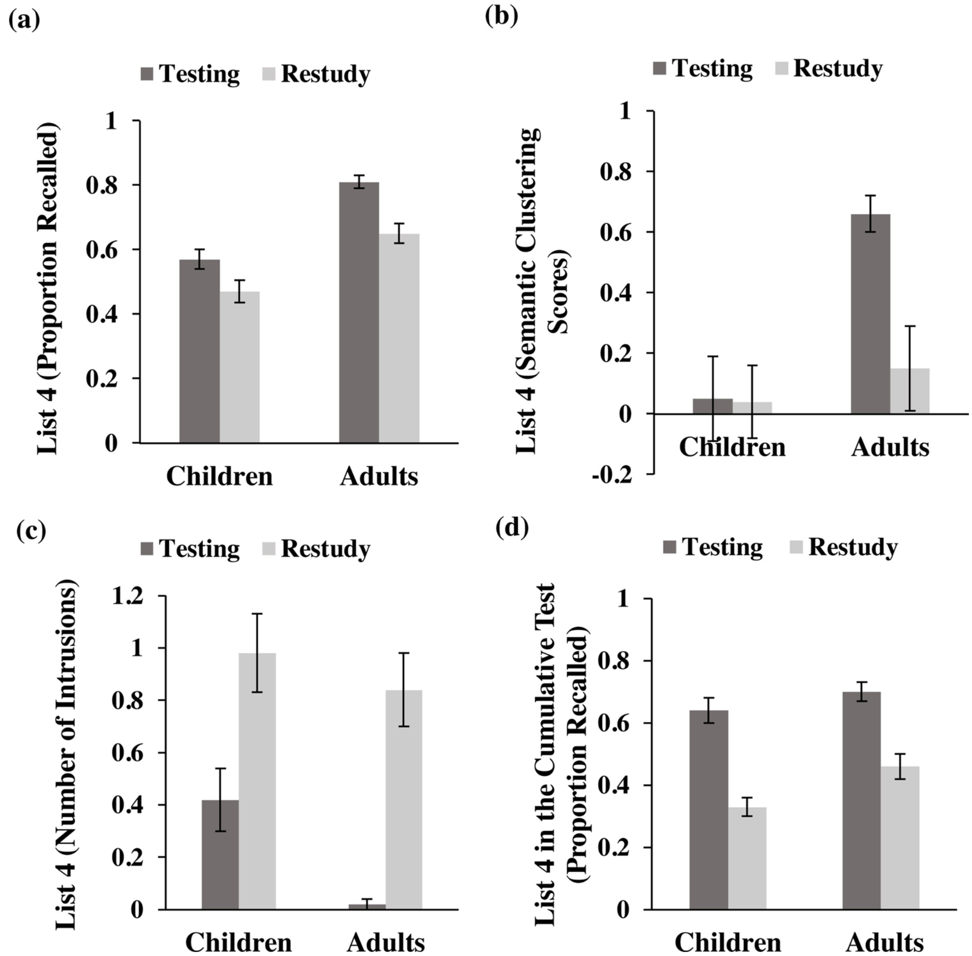

Fig. 2. (a) Correct recall in the List 4 interim test as a function of age (children *vs*. adults) and interim task (test *vs*. restudy). (b) Semantic clustering scores in List 4 interim test as a function of age (children *vs*. adults) and interim task (test *vs*. restudy). (c) Mean of prior-list intrusions from Lists 1–3 during recall of List 4 words as a function of age (children *vs*. adults) and interim task (test *vs*. restudy). (d) Correct recall in the List 4 cumulative test as a function of age (children *vs*. adults) and interim task (test *vs*. restudy). Error bars represent ± 1 standard error.

Table 1
Proportions (*SD*s) of List 1-4 interim test recall.

List	Interim task	Children	Adults
List 1	Test	0.56 (0.19)	0.77 (0.15)
	Restudy		
List 2	Test	0.54 (0.16)	0.79 (0.15)
	Restudy		
List 3	Test	0.58 (0.16)	0.77 (0.12)
	Restudy		
List 4	Test	0.57 (0.20)	0.81 (0.14)
	Restudy	0.47 (0.24)	0.65 (0.17)

following, participants in the test condition were asked to orally recall as many words as they could from the just-studied list in any order they wished within 30 s. By contrast, those in the restudy condition restudied the words for Lists 1–3 one more time, in a new random order for words in each list. But they were tested on List 4. Thus, the procedure for List 4 was the same between the two conditions, but different for Lists 1–3. Participants' verbal responses were recorded by the experimenter.

Following the completion of List 4, participants were dismissed and invited to come back 24 h later to take part in the cumulative

Table 2
Semantic clustering scores (*SDs*) in List 1-4 interim tests.

List	Interim task	Children	Adults
List 1	Test Restudy	0.03 (0.92)	0.19 (0.81)
List 2	Test Restudy	0.06 (0.89)	0.51 (0.60)
List 3	Test Restudy	0.04 (1.04)	0.52 (0.71)
List 4	Test Restudy	0.05 (0.95) 0.04 (0.80)	0.66 (0.42) 0.15 (0.94)

test, during which they recalled as many words as they could from all lists in 5 min. No feedback was provided in any interim or cumulative test.

3. Results

Note that when non-significant results were observed, we conducted Bayesian analyses to evaluate whether and to what extent the non-significant results supported the null hypotheses.

3.1. List 4 interim test results

3.1.1. Correct recall

As shown in Fig. 2a (also see Table 1), a 2 (age: children *vs.* adults) × 2 (interim task: test *vs.* restudy) analysis of variance (ANOVA) revealed a main effect of age on List 4 interim test recall, $F(1, 176) = 56.13$, $p < .001$, $\eta_p^2 = .24$. Correct recall increased with age (children: $M = 0.52$, $SD = 0.22$; adults: $M = 0.73$, $SD = 0.18$). Pairwise comparisons showed significant differences between children and adults: $p < .001$, $d = 1.13$.

The main effect of interim task was significant, $F(1, 176) = 22.62$, $p < .001$, $\eta_p^2 = .11$, and List 4 interim test recall was higher in the test ($M = 0.69$, $SD = 0.21$) than in the restudy condition ($M = 0.56$, $SD = 0.23$), reflecting the FTE.

The interaction between the two factors was non-significant, $F(1, 176) = 0.96$, $p = .33$, $\eta_p^2 = .01$, $BF_{01} = 3.19$, indicating that the observed data supported the null hypothesis over the alternative hypothesis. Pairwise comparisons showed that the FTE was robust in both children, $p = .008$, $d = 0.57$, and adults, $p < .001$, $d = 0.86$, replicating the main findings from Aslan and Bäuml (2016).

3.1.2. Semantic clustering

Adjusted-ratio-of-clustering (ARC, Roenker et al., 1971) quantifies the likelihood that semantically related items follow each other during recall (*i.e.*, semantic clustering), with positive ARC scores indicating above chance clustering, 0 indicating chance level clustering, and negative scores indicating below chance clustering. In this analysis, we followed Chan, Manley et al. (2018) to substitute an undefined ARC score with 0. This occurs when only one item is recalled from each category or when all of the recalled items are from the same category.

As shown in Fig. 2b (also see Table 2), a 2 (age: children *vs.* adults) × 2 (interim task: test *vs.* restudy) ANOVA showed a main effect of age, $F(1, 173) = 8.75$, $p = .004$, $\eta_p^2 = .05$, with children showing inferior semantic clustering than adults: $p = .003$, $d = 0.45$. The main effect of interim task was significant, $F(1, 173) = 4.58$, $p = .03$, $\eta_p^2 = .03$, reflecting superior clustering in the test ($M = 0.36$, $SD = 0.79$) than in the restudy condition ($M = 0.09$, $SD = 0.87$). There was a significant interaction between age and interim task, $F(1, 173) = 4.22$, $p = .04$, $\eta_p^2 = .02$.

Pairwise comparisons showed that, in comparison with restudy, test significantly increased semantic clustering scores for adults, $p = .003$, $d = 0.63$, but not for children, $p = .95$, $d = 0.01$, $BF_{01} = 4.46$. It reflects that the beneficial effect of interim testing on semantic clustering was determined by age. In addition, adults' ARC scores were significantly greater than the chance level, $p < .001$, $d = 1.05$, while ARC scores of children were near to the chance level, $p = .73$, $d = 0.10$, $BF_{01} = 7.59$.

3.1.3. Prior-list intrusions

As shown in Fig. 2c, a 2 (age: children *vs.* adults) × 2 (interim task: test *vs.* restudy) ANOVA revealed a main effect of interim task on the number of prior-list intrusions in the List 4 interim test, $F(1, 176) = 34.53$, $p < .001$, $\eta_p^2 = .16$, such that intrusions occurred less frequently in the test condition ($M = 0.22$, $SD = 0.61$) than in the restudy condition ($M = 0.91$, $SD = 0.94$), $p < .001$, $d = 0.88$. The main effect of age was significant, $F(1, 176) = 5.17$, $p = .024$, $\eta_p^2 = .03$. Intrusions decreased as a function of age (children: $M = 0.70$, $SD = 0.94$; adults: $M = 0.43$, $SD = 0.77$), $p = .024$, $d = 0.34$. The interaction between age and interim task was non-significant, $F(1, 176) = 1.29$, $p = .26$, $\eta_p^2 = .01$, $BF_{01} = 2.56$.

Table 3
Pearson correlations among variables for children and adults.

	Children				Adults			
	1.	2.	3.	4.	1.	2.	3.	4.
1. Interim task	1				1			
2. List 4 interim test recall	0.24*	1			0.46 ***	1		
3. Semantic clustering scores	0.01	0.11	1		0.33**	0.40 ***	1	
4. Prior-list intrusions	−0.30**	−0.38***	0.05	1	−0.54***	−0.52***	−0.23*	1

Note: * $p < .05$; ** $p < .01$, *** $p < .001$.

Table 4
Results from a moderated, multiple mediation analysis.

The moderate, multiple mediation model	β	95 % CI
Interim task – (Semantic clustering scores, Prior-list intrusions) – List 4 interim test recall		
Total effect for children	0.101	[0.033, 0.170]
Total effect for adults	0.159	[0.096, 0.219]
Direct effect	0.034	[−0.024,0.093]
Indirect effect through Semantic clustering scores for children	0.001	[−0.025, 0.023]
Indirect effect through Semantic clustering scores for adults	0.032	[0.014, 0.053]
Difference in indirect effects through Semantic clustering scores between children and adults	−0.031	[−0.066, −0.002]
Indirect effect through Prior-list intrusions for children	0.067	[0.022, 0.115]
Indirect effect through Prior-list intrusions for adults	0.093	[0.061, 0.129]
Difference in indirect effect through Prior-list intrusions between children and adults	−0.026	[−0.080, 0.028]
Total indirect effect for children	0.068	[0.019, 0.118]
Total indirect effect for adults	0.125	[0.086, 0.166]
Difference in indirect effects between Prior-list intrusions and Semantic clustering score for children	0.067	[0.016, 0.122]
Difference in indirect effects between Prior-list intrusions and Semantic clustering score for adults	0.061	[0.024, 0.101]

3.2. List 4 recall in the cumulative test

Of particular interest was the List 4 cumulative test recall. As shown in Fig. 2d (also see Appendix A. Table A2), a 2 (age group: children *vs.* adults) × 2 (interim task: test *vs.* restudy) between-subjects ANOVA, with List 4 cumulative test recall as the dependent variable, revealed a main effect of age, $F(1, 176) = 6.61$, $p = .01$, $\eta_p^2 = .04$. That is, correct recall increased with age (children: $M = 0.49$, $SD = 0.29$; adults: $M = 0.58$, $SD = 0.28$), $p = .01$, $d = 0.38$. And the main effect of interim task was significant, $F(1, 176) = 57.96$, $p < .001$, $\eta_p^2 = .25$, reflecting higher recall in the test ($M = 0.67$, $SD = 0.24$) than in the restudy condition ($M = 0.39$, $SD = 0.26$), $p < .001$, $d = 1.13$. Moreover, the interaction between age and interim task was non-significant, $F(1, 176) = 1.12$, $p = .29$, $\eta_p^2 = .01$, $BF_{01} = 2.92$.

The results of List 1–3 cumulative test recall are reported in the Appendix A.

3.3. Interim test recall across lists

This section focuses on changes in interim test recall across Lists 1–4. Because participants in the restudy condition did not take interim tests on Lists 1–3, their data were not included in the analyses below. Recall performance across lists for participants in the test condition was analyzed using a 4 (Lists 1–4) × 2 (children *vs.* adults) mixed ANOVA.

As shown in Table 1, recall performance across lists increased with age, $F(1, 88) = 79.65$, $p < .001$, $\eta_p^2 = .48$. There were no main effect of list, $F(3, 264) = 1.15$, $p = .33$, $\eta_p^2 = .01$, $BF_{01} = 18.70$, and no main effect of interaction between list and age, $F(3, 264) = 0.81$, $p = .49$, $\eta_p^2 = .01$, $BF_{01} = 13.51$.

ARC scores across lists were analyzed using a 4 (Lists 1–4) × 2 (children *vs.* adults) mixed ANOVA. As shown in Table 2, ARC scores were greater for adults than those for children, $F(1, 87) = 22.40$, $p < .001$, $\eta_p^2 = .21$. There was no main effect of list, $F(3, 261) = 1.46$, $p = .23$, $\eta_p^2 = .02$, $BF_{01} = 10.96$, and no significant interaction between list and age, $F(3, 261) = 1.37$, $p = .25$, $\eta_p^2 = .02$, $BF_{01} = 6.06$.

3.4. Age difference in the roles of semantic processing strategy-change and release-from-PI in the FTE

The above results suggested that release-from-PI may contribute to both children's and adults' FTEs, whereas semantic processing strategy-change may only contribute to adults' but not children's FTE. As shown in Table 3, the correlation results imply that interim task (test = 1; restudy = 0) was positively correlated with List 4 interim test recall for both children and adults. In addition, interim testing enhanced clustering scores for adults but not for children. By contrast, interim testing reduced prior-list intrusions for both children and adults. Furthermore, List 4 interim test recall was positively correlated with semantic-clustering scores and was negatively correlated with prior-list intrusions.

To determine whether semantic processing strategy-change and release-from-PI contributed to the observed FTE, and whether the mediations were moderated by age, a moderated, multiple mediation analysis was conducted, in which interim task was treated as the independent variable (coded as a dummy variable: test = 1; restudy = 0), List 4 interim test recall as the dependent variable, semantic clustering scores and prior-list intrusions as the mediators, and age as the moderator (children = 1; adults = 0).

This moderated, multiple mediation model was fitted using the free, open-source SEM-package *lavaan* (Rosseel, 2012) in *R*. We generated 95 % bootstrap bias corrected confidence intervals (CI) for the indirect effects conditioned by age with 5000 bootstrap samples. Bayesian tests of informative hypotheses about parameters in the model were also conducted in *R*, using the *bain* package (Gu, Hoijtink, Mulder, & Rosseel, 2019; Van Lissa et al., 2020). The BF_{01} was reported to supplement the null effects. Specifically, the null-hypothesis was evaluated by constraining the parameters to be equal to zero and the BF_{01} was computed by comparing it with the alternative unconstrained hypothesis (for details, see Van Lissa et al., 2020).

The detailed results are shown in Table 4. The indirect effect *via* semantic processing strategy-change for children was 0.001 [−0.025, 0.023], $BF_{01} = 8.36$. By contrast, the indirect effect for adults was 0.032 [0.014, 0.053]. These results suggest a mediating role of sematic processing strategy-change in the FTE for adults but not for children. The difference in indirect effects between children and adults was −0.031 [−0.066, −0.002], suggesting a moderating effect of age on this indirect effect.

The indirect effect through release-from-PI (indexed by prior-list intrusions) was significant for both children, 0.067 [0.022, 0.115], and adults, 0.093 [0.061, 0.129], reflecting a mediating role of release-from-PI in both children's and adults' FTE. The difference in indirect effects between children and adults was −0.026 [−0.080, 0.028], $BF_{01} = 5.91$, suggesting a minimal moderating effect of age on this indirect effect. Put differently, release-from-PI played a roughly equally important role in children's and adults' FTE.

The above findings imply that only release-from-PI contributed significantly to children's FTE, and this mechanism successfully accounted for about 66.34 % (0.067 ÷ 0.101 × 100 %) of the total effect of interim task on List 4 interim test recall (*i.e.*, FTE).

These findings also imply that both semantic processing strategy-change and release-from-PI contributed significantly to adults' FTE. The indirect effect through semantic processing strategy-change accounted for about 20.13 % (0.032 ÷ 0.159 × 100 %) of adults' FTE, and release-from-PI accounted for about 58.49 % (0.093 ÷ 0.159 × 100 %) of adults' FTE. Furthermore, for adults, the difference in indirect effects between prior-list intrusions and semantic clustering scores was significant, difference = 0.061 [0.024, 0.101], suggesting that release-from-PI played a more important role than semantic processing strategy-change in adults' FTE.

4. Discussion

Although it has been repeatedly established that the FTE is a robust phenomenon and of practical significance, its cognitive underpinnings remain elusive. Moreover, the majority of previous studies explored the FTE in adults, but little is known about whether the FTE exists in children, and whether there is an age difference in mechanisms underlying the FTE. The current study filled these gaps by testing the FTE in children and adults and employing a moderated, multiple mediation model to explore the developmental trajectory of mechanisms underlying the FTE.

First, we conceptually replicated prior work (Aslan & Bäuml, 2016) by observing that for both third-grade children and adults, interim testing enhanced List 4 interim test recall and prevented the build-up of PI across lists (in line with Hypothesis 1).

Second, the documented results shed new light on the developmental trajectory of mechanisms underlying the FTE. Specifically, age (children *vs.* adults) moderated the first stage of the indirect effect through semantic clustering, with adults' but not children's FTE showing a significant mediation effect *via* semantic clustering scores (consistent with Hypothesis 2). More importantly, age failed to significantly moderate the indirect effect *via* prior list intrusions, with both children and adults showing this mediation effect (in support of Hypothesis 3). Overall, these results reflect that both the semantic processing strategy-change mechanism (accounting for 20.13 % of adults' FTE) and the release-from-PI mechanism (accounting for 58.49 % of adults' FTE) contributed significantly to adults' FTE, while only the release-from-PI mechanism (accounting for 66.34 % of children's FTE) contributed significantly to children's the FTE.

Third, the relative contributions of the two mechanisms were compared. For adults, release-from-PI played a more important role than semantic processing strategy-change. It should be noted that the current study is the first to (1) provide mediation evidence supporting the release-from-PI account and the semantic processing strategy-change account, and (2) offer important and novel insights into the developmental trajectory of the mechanisms underlying the FTE. In particular, compared to the semantic processing strategy-change mechanism, children's FTE is primarily driven by the release-from-PI mechanism but is unlikely to be attributed to the semantic processing strategy-change mechanism. By contrast, both the release-from-PI mechanism and the semantic processing strategy-change mechanism are responsible for adults' FTE. Although both mechanisms are responsible for adults' FTE, the release-from-PI mechanism plays a dominant role whereas the semantic processing strategy-change mechanism plays a less important role.

Prior research proposed a semantic processing strategy-change account (Chan, Manley et al., 2018; Chan et al., 2020). The current study further clarified that semantic processing strategy-change tends to play a causal role in the FTE. That is, interim testing enhances new learning by changing individuals' semantic processing strategies during subsequent encoding and retrieval of new information. Specifically, during the interim test, participants might use the recalled items as a cue to facilitate retrieval of other items from the same category (Carpenter, 2011; Pyc & Rawson, 2010). This process might alter participants' encoding strategy for the upcoming study lists, as they will tend to organize the words into different semantic categories to facilitate encoding, during which the semantic processing is enhanced not only among related items in the study list, but also among those spontaneously retrieved from prior lists (study-phase retrieval, *e.g.*, Wahlheim, 2015). Meanwhile, prior interim tests inform participants of the test format to expect (Cho et al., 2017) and the type of retrieval cues. Thus, participants increasingly utilized semantic processing strategies to aid encoding and retrieving new information. The result from the current study (*i.e.*, ARC scores mediated the positive relationship between interim task and List 4 interim test recall) is consistent with the above assumptions.

Critically, the current findings show that the mediation effect was absent in children's FTE. A possible explanation is that children failed to spontaneously invoke appropriate strategies in a given situation (Brown & DeLoache, 1978; Moely, Olson, Halwes, & Flavell, 1969) and were not invoked to apply those strategies by interim tests. In particular, children in the current study were incapable of using effective semantic processing strategies (as reflected by their poor ARC scores). Besides, the results showed that children's ARC scores did not vary across lists in the test condition, suggesting that they were inferior to adults in reasoning about the subsequent test format, and failed to modify processing strategies to potentiate recall performance. These results also suggest that interim testing is insufficient to promote children's semantic clustering. They might require more explicit prompts from the experimenter (Bjorklund et al., 1977) or stronger semantic associations among list items (Corsale & Ornstein, 1980; Schneider, 1986) to display semantic organization.

Apart from this, the results might result from the fact that adopting semantic processing strategies requires large amounts of mental effort, and children are reluctant to employ such demanding strategies. For instance, the greater the required level of clustering, the greater the processing demands on working memory, and thus, less capacity is left for encoding and storage. Adults, by comparison with children, have more cognitive resources, they, therefore, can execute a strategy more efficiently than children (Bjorklund & Harnishfeger, 1987; DeMarie-Dreblow & Miller, 1988; Guttentag, 1984). Indeed, prior research demonstrated that, although testing enhanced recall of previously studied information by comparison with restudying (the BTE) for children aged 7 and 9, it failed to enhance children's semantic clustering during free recall (Lipowski et al., 2014).

It should be noted that the release-from-PI account (Szpunar et al., 2008; Yang et al., 2018) also accounts for the documented FTE. The findings showed that immediate testing of studied lists reduced prior-list intrusions. More importantly, prior-list intrusions mediated the relation between interim task and List 4 interim recall for both children and adults, suggesting that interim testing causes context changes, which in turn reduces PI and enhances learning of new information. Likewise, Ma, Li, Li, and Zhou (2020) found that episodic context reinstatement promoted memory retention in children and adults.

No difference in the mediation effect *via* prior-list intrusions was found between adults' and children's FTEs, suggesting that, compared to the role of semantic-processing strategy-change which is determined by age, release-from-PI is functional and plays a roughly equally important role in both children's and adults' FTEs. A possible explanation is that there is a dissociation between children's and adults' learning in terms of strategy use (strategy-change based) but not mental context change (release-from-PI based). Strategy change is indirectly induced by interim testing, such that participants may firstly reason about the nature of later tests (children are poor at such reasoning) and then adopt more effective strategies (children are incapable of using effective semantic strategies) that benefit subsequent new learning. While context change is directly invoked by interim testing (Szpunar et al., 2008), which does not necessitate this reasoning process. Therefore, both children and adults benefit from test-induced context change.

Taken together, the current study suggests that two mechanisms underlie the FTE's development trajectory. Only the release-from-PI mechanism (but not the semantic processing strategy-change) was operative in children's FTE. While both the release-from-PI and the semantic processing strategy-change mechanisms were responsible for adults' FTE, and the former played a more important role than the latter. Thus, the release-from-PI mechanism developed earlier than the semantic processing strategy-change mechanism.

After a 24-h delay, participants in the test condition still substantially outperformed those in the restudy condition as demonstrated by the superior Lists 1–3 and List 4 cumulative test recall in the test condition. First, the difference observed in List 4 cumulative test recall between the test and restudy conditions can be partly attributed to more attention and effort (induced by testing Lists 1–3) directed to learning Lists 4, and/or due to additional exposure, that is, participants in the test condition recalled significantly more List 4 words than did those in the restudy condition on Day 1, thereby providing themselves with an additional exposure of these words.

Second, the difference observed in Lists 1–3 cumulative test recall can be attributed to a combination of a standard BTE and more attention and effort directed to learning Lists 2–3 (Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Szpunar, Khan, & Schacter, 2013). For example, children in the test condition recalled many List 3 words and those in the restudy condition restudied all List 3 words. Thus, the additional exposure of the List 3 words between the two conditions may have little difference. However, a combination of a standard BTE and more attention and effort directed to learning List 3 may result in the superior List 3 cumulative test recall. Because children hadn't taken the test before they study List 1, and they had taken one, two, or three tests before they study List 2, List 3, or List 4, thus, testing may have motivated children to put forth more effort when they studied List 3 and List 4 than List 1 and List 2. It may lead to the result that the difference in recall between the two conditions was three times as large for the children on the cumulative recall test for Lists 3 and 4 words in comparison with their recall for Lists 1 and 2 words.

Third, Table 2 showed that adults in the test condition recalled many more words than did children in that condition for all lists on Day 1 and the recall of Lists 1 and 2 on the cumulative test on Day 2, but the children in the test condition recalled essentially as many Lists 3 and 4 words on the cumulative test as did the adults. One possible reason is that participants in the test condition were asked to recall words from the specific list that they had studied on the interim test on Day 1; given that children can experience greater difficulty than adults in combating proactive interference from the previously studied list, they recalled fewer words than adults in that condition for all lists. However, on Day 2, participants were asked to recall words from all lists on the cumulative test, thus they were not limited to recall words from a specific list. In this case, children were less likely to be affected by proactive interference, particularly in List 3 and 4 recall (the buildup of proactive interference was caused by the extended study sessions, which led to the proactive interference in List 3 and 4 was more than List 1 and 2). Accordingly, children recalled essentially as many words from List 3 and 4 on the cumulative test as did the adults.

4.1. Limitations and future directions

Although the current study observed that semantic processing strategy-change underlies the FTE in adults, the results provide little

Table A1
Three categories and eight items in each category used as stimuli in Chinese and English translation.

水果Fruits	身体部分Body parts	动物Animals
芒果Mango	手指Fingers	鲨鱼Shark
樱桃Cherry	脑袋Head	乌龟Turtle
蓝莓Blueberry	耳朵Ear	斑马Zebra
荔枝Lychee	大腿Thigh	蝴蝶Butterfly
葡萄Grapes	牙齿Teeth	蚂蚁Ant
橘子Orange	眉毛Eyebrows	奶牛Cow
菠萝 Pineapple	肚子Belly	老鼠 Mouse
榴莲 Durian	肩膀 Shoulder	狐狸 Fox

Table A2
Proportions (*SDs*) of List 1–4 cumulative test recall.

List	Interim task	Children	Adults
List 1	Test	0.37 (0.24)	0.63 (0.19)
	Restudy	0.26 (0.19)	0.40 (0.25)
List 2	Test	0.30 (0.21)	0.56 (0.25)
	Restudy	0.19 (0.15)	0.34 (0.24)
List 3	Test	0.61 (0.23)	0.64 (0.25)
	Restudy	0.31 (0.18)	0.42 (0.21)
List 4	Test	0.64 (0.26)	0.70 (0.22)
	Restudy	0.33 (0.21)	0.46 (0.28)

information about the roles of other strategies in the FTE, such as the role of temporal processing strategy-change. Nairne, Cogdill, and Lehman (2017) noted that temporal organization or clustering indexes the degree to which items studied in neighboring serial positions in a list tend to be reported together during the recall output sequence. Considering that testing increases temporal clustering across three successive lists (Zacks, Hasher, Alba, Sanft, & Rose, 1984), and the ability to form temporal associations increases dramatically between the ages of 9 and 10 (Guillery-Girard et al., 2013), both children and adults may succeed to use this strategy in the test condition. Future studies should clarify this issue by using unrelated word lists and comparing temporal clustering scores between the test and restudy conditions (Polyn, Norman, & Kahana, 2009).

After completing the List 4 interim test, we did not ask participants to report strategies they used. Interim testing may promote participants to use more elaborate encoding strategies such as sentence generation and mental imagery (Soderstrom & Bjork, 2014). Generally, children become increasingly effective in implementing these mnemonic strategies with development. At what age the strategy change mechanism starts to work is unknown. Future research should go beyond the current work to employ a wider range of age groups and include measures of different strategies.

The fact that semantic processing strategy-change fails to account for the FTE in children does not necessarily imply that the release-from-PI account is the sole contributor in this case. Many possible explanations have been proposed for this effect (for a review, see Yang et al., 2018). Similarly, a single mechanism (*i.e.*, semantic processing strategy-change) may not be sufficient to explain the age difference in the FTE, whereas other mechanisms may work together to account for the age difference. Therefore, a promising direction for future research is to explore how different mechanisms work in conjunction to account for the age difference in the FTE.

4.2. Conclusion

To summarize, the FTE generalizes to third-grade children and adults. Release-from-PI plays an equally important role in both children's and adults' FTE, whereas semantic processing strategy-change is only responsible for adults' but not children's FTE. Besides, release-from-PI plays a more important role in adults' FTE than semantic processing strategy-change. The release-from-PI mechanism develops earlier than the semantic processing strategy-change mechanism.

Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgement

This study was supported by major project grants from the National Social Science Foundation of China (No. 14ZDB160).

Appendix A

Lists 1–3 cumulative test recall were analyzed using a 2 (interim task: test *vs.* restudy) × 2 (children *vs.* adults) between-subjects ANOVA. A main effect of age was observed, $F(1, 176) = 47.79$, $p < .001$, $\eta_p^2 = .21$. That is, correct recall for adults ($M = 0.50$, $SD = 0.20$) was greater than that for children ($M = 0.34$, $SD = 0.17$), $p = .02$, $d = 1.09$. And there was a main effect of interim task, $F(1, 176) = 74.45$, $p < .001$, $\eta_p^2 = .30$, with superior recall performance in the test condition ($M = 0.52$, $SD = 0.19$) by comparison with that in the restudy condition ($M = 0.32$, $SD = 0.15$), $p < .001$, $d = 1.22$. The interaction between age and interim task was non-significant, $F(1, 176) = 0.93$, $p = .34$, $\eta_p^2 = .01$, $BF_{01} = 3.13$.

Appendix B. Supplementary material

The data report is publicly available at https://osf.io/zpnvc/.

References

Aslan, A., & Bäuml, K. H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science, 19*, 6. https://doi.org/10.1111/desc.12340

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

Bjorklund, D. F., & Harnishfeger, K. K. (1987). Developmental differences in the mental effort requirements for the use of an organizational strategy in free recall. *Journal of Experimental Child Psychology, 44*(1), 109–125. https://doi.org/10.1016/0022-0965(87)90025-7

Bjorklund, D. F., Ornstein, P. A., & Haig, J. R. (1977). Developmental differences in organization and recall: Training in the use of organizational techniques. *Developmental Psychology, 13*(3), 175–183. https://doi.org/10.1037/0012-1649.13.3.175

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology, 49*(2), 229–240. https://doi.org/10.1080/00221309.1953.9710088

Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. (1958). Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports, 4*(1), 39–44. https://doi.org/10.2466/pr0.1958.4.g.39

Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. *Children's thinking: What develops* (pp. 3–35).

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547–1552. https://doi.org/10.1037/a0024140

Chan, J. C. K., Manley, K. D., & Ahn, D. (2020). Does retrieval potentiate new learning when retrieval stops but new learning continues? *Journal of Memory and Language, 115*, Article 104150. https://doi.org/10.1016/j.jml.2020.104150

Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language, 102*, 83–96. https://doi.org/10.1016/j.jml.2018.05.007

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

Cho, K. W., Neely, J. H., Crocco, 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. https://doi.org/10.1080/17470218.2016.1175485

Corsale, K., & Ornstein, P. A. (1980). Developmental changes in children's use of semantic information in recall. *Journal of Experimental Child Psychology, 30*(2), 231–245. https://doi.org/10.1016/0022-0965(80)90060-0

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

DeMarie-Dreblow, D., & Miller, P. H. (1988). The development of children's strategies for selective attention: Evidence for a transitional period. *Child Development, 59*(6), 1504–1513. https://doi.org/10.2307/1130665

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Gaskill, P. J., & Murphy, P.k. (2004). Effects of a memory strategy on second-graders' performance and self-efficacy. *Contemporary Educational Psychology, 29*(1), 27–49. https://doi.org/10.1016/s0361-476x(03)00008-0

Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation, 89*, 1526–1553. https://doi.org/10.1080/00949655.2019.159057

Guillery-Girard, B., Martins, S., Deshayes, S., Hertz-Pannier, L., Chiron, C., Jambaqué, I., et al. (2013). Developmental trajectories of associative memory from childhood to adulthood: A behavioral and neuroimaging study. *Frontiers in Behavioral Neuroscience, 7*, 126. https://doi.org/10.3389/fnbeh.2013.00126

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology, 37*(1), 92–106. https://doi.org/10.1016/0022-0965(84)90060-2

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation, 61*, 237–284. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50*(4), 994–1000. https://doi.org/10.1037/a0035202

Ma, X., Li, T., Li, Z., & Zhou, A. (2020). Episodic context reinstatement promotes memory retention in older but not younger elementary schoolchildren. *British Journal of Developmental Psychology, 38*(2), 304–318. https://doi.org/10.1111/bjdp.12321

Moely, B. E., Olson, F. A., Halwes, T. G., & Flavell, J. H. (1969). Production deficiency in young children's clustered recall. *Developmental Psychology, 1*(1), 26–34. https://doi.org/10.1037/h0026804

Nairne, J. S., Cogdill, M., & Lehman, M. (2017). Adaptive memory: Temporal, semantic, and rating-based clustering following survival processing. *Journal of Memory and Language, 93*, 304–314. https://doi.org/10.1016/j.jml.2016.10.009

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

Pastötter, B., Engel, M., & Frings, C. (2018). The forward effect of testing: Behavioral evidence for the reset-of-encoding hypothesis using serial position analysis. *Frontiers in Psychology, 9*, 1197. https://doi.org/10.31234/osf.io/8wzmt

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. https://doi.org/10.1037/a0021801

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129–156. https://doi.org/10.1037/a0014420

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002). https://doi.org/10.1126/science.1191465, 335-335.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*(1), 45–48. https://doi.org/10.1037/h0031355

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36,. http://www.jstatsoft.org/v48/i02/.

Schleepen, T. M. J., & Jonkman, L. M. (2012). Children's use of semantic organizational strategies is mediated by working memory capacity. *Cognitive Development, 27*(3), 255–269. https://doi.org/10.1016/j.cogdev.2012.03.003

Schneider, W. (1986). The role of conceptual knowledge and metamemory in the development of organizational processes in memory. *Journal of Experimental Child Psychology, 42*(2), 218–236. https://doi.org/10.1016/0022-0965(86)90024-X

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty*. Psychology Press.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99–115. https://doi.org/10.1016/j.jml.2014.03.003

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). "Testing during study insulates against the buildup of proactive interference": Correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392–1399. https://doi.org/10.1037/a0013082

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. https://doi.org/10.1073/pnas.1221764110

Van Lissa, C. J., Gu, X., Mulder, J., Rosseel, Y., Van Zundert, C., & Hoijtink, H. (2020). Teacher's corner: Evaluating informative hypotheses using the Bayes factor in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–10. https://doi.org/10.1080/10705511.2020.1745644

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

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. https://doi.org/10.1037/a0036164.supp

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

Wimmer, H., & Tornquist, K. (1980). The role of metamemory and metamemory activation in the development of mnemonic performance. *International Journal of Behavioral Development, 3*(1), 71–81. https://doi.org/10.1177/016502548000300107

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(3), 485–492. https://doi.org/10.1037/xlm0000449

Yang, C., Chew, S. J., 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. https://doi.org/10.1037/xap0000122

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). https://doi.org/10.1038/s41539-018-0024-y

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2020). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*. Advance online publication.

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000278. Advance online publication.

Zacks, R. T., Hasher, L., Alba, J. W., Sanft, H., & Rose, K. C. (1984). Is temporal order encoded automatically? *Memory & Cognition, 12*(4), 387–394. https://doi.org/10.3758/bf03198299

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38*(8), 995–1008. https://doi.org/10.3758/MC.38.8.995