European Journal of Developmental Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pedp20

Developmental trajectory of the forward testing effect: The role of reset-of-encoding

Xixi Dang, Chunliang Yang, Mengying Che, Yinghe Chen & Xiao Yu

To cite this article: Xixi Dang, Chunliang Yang, Mengying Che, Yinghe Chen & Xiao Yu (2021): Developmental trajectory of the forward testing effect: The role of reset-of-encoding, European Journal of Developmental Psychology, DOI: 10.1080/17405629.2021.1986386

To link to this article: https://doi.org/10.1080/17405629.2021.1986386

Published online: 11 Oct 2021.

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY
https://doi.org/10.1080/17405629.2021.1986386

Developmental trajectory of the forward testing effect: The role of reset-of-encoding

Xixi Dang[a], Chunliang Yang[b], Mengying Che[b], Yinghe Chen[b] and Xiao Yu[c]

[a]Department of Psychology, Zhejiang Sci-Tech University, Hangzhou, China; [b]School of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China; [c]Department of Psychology, School of Humanities and Social Sciences, Beijing Forestry University, Beijing, China

ABSTRACT

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 the developmental trajectory of the FTE and whether the reset-of-encoding process contributes to the FTE. Younger children, older children, and adults were instructed to study four lists of unrelated words, then either restudied or were tested following studying each of Lists 1–3, and took an interim test on List 4. Results demonstrated that interim testing on Lists 1–3 enhanced learning of List 4 for younger children, older children, and adults. Importantly, this enhancement varies with items' serial list position in both younger children and older children. Early List 4 items at list primacy positions benefited more from interim testing than later ones at non-primacy positions. Overall, this is the first study demonstrating that (1) the FTE generalizes to younger children, older children, and adults, suggesting the FTE emerges quite early in human life; (2) the reset-of-encoding process contributes to the FTE in both younger and older children.

ARTICLE HISTORY Received 24 December 2020; Accepted 22 September 2021

KEYWORDS Forward testing effect; developmental trends; reset-of-encoding

Introduction

The forward effect of testing (FTE) is the ironic finding that interim testing of previously studied information potentiates subsequent learning of new materials (for reviews, see Pastötter & Bäuml, 2014; Yang et al., 2018). In a standard FTE experiment, participants study several lists of items in anticipation of a final cumulative recall test. Participants are asked to either retrieve what they remember from the just-studied set of materials (interim test condition), complete an unrelated task between studying

CONTACT Yinghe Chen chenyinghe@bnu.edu.cn School of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing 100875, China

each set of material (no test control), or restudy the just-studied set of materials (restudy control) before they study the target (new) materials. Critically, all participants are tested on the target set after studying. The classic finding, derived from this research paradigm, is that interim testing of previously studied materials potentiates subsequent learning of new materials (e.g., with more items correctly recalled from the target new set), relative to the control counterparts.

The FTE has been proven as a robust phenomenon by using a variety of study materials in different experimental settings (Chan, Meissner, et al., 2018; Pastötter & Bäuml, 2014; Yang et al., 2018), and this effect generalizes across different populations, including college students with varying levels of working memory capacity (Pastötter & Frings, 2019) and test anxiety (Yang et al., 2020), older adults (Pastötter & Bäuml, 2019), and individuals with severe traumatic brain injury (Pastötter et al., 2013). However, the majority of previous studies explored the FTE in adults, and it remains largely unknown whether the FTE generalizes to young children and what its developmental trajectory is. Thus, the current study aims to determine the developmental trend of the FTE, and unravel the mechanisms underlying the effect. Specifically, the current study focuses on a reset-of-encoding account which ascribes the FTE to a reset of the encoding process (Pastötter et al., 2018, 2011).

FTE in children

The majority of previous studies on the FTE have been carried out with college students as participants, with its generalizability to young children and its developmental trajectory largely underexplored. To our knowledge, only one study has evaluated the FTE in children (Aslan & Bäuml, 2016). In this study, younger (average age = 6.7 years) and older (average age = 8.8 years) elementary school children and adults studied four lists of items in anticipation of a final cumulative recall test. Following the presentation of each of the first three lists, participants either completed a cued recall test on the respective list, or restudied that list. All participants engaged in a free recall test on the last target list. Results demonstrated that the FTE generalizes to older children and adults but not to young children.

Obviously one study is insufficient to draw a firm conclusion. Indeed, other research (Otgaar et al., 2019) suggests that even younger children's learning can benefit from interim testing. In Otgaar et al.'s (2019)

Experiment 1, young children (mean age = 7.32) viewed a lecture video, and then half of them were interviewed while the other half were not interviewed. After that, children received misinformation and attended a final memory test. This study found that retrieval practice increases children's suggestibility to subsequently presented misleading information. As suggested by Otgaar et al.'s (2019), retrieval practice can enhance new learning by reducing inattention (Pastötter et al., 2011; Szpunar et al., 2013), such that children who had encountered an earlier retrieval practice (i.e., the interview) were more likely to sustain their attention for the subsequent misleading information. This may make the misleading information learning as effective as prior video learning, inducing suggestibility to the misleading information in young children. This rest-of-encoding process (see the next section for more details) may be responsible for both the finding of Otgaar et al. (2019) and the FTE, suggesting the FTE might generalize to young children.

In sum, there is a procedural difference between Aslan and Bäuml (2016) study (mismatched test format between prior nontarget list tests and target list test) and the standard FTE experiment (matched test format between prior nontarget list tests and target list test). In addition, Otgaar et al. (2019) provided inconsistent findings suggesting that the FTE may exist in young children. It is therefore difficult to draw any firm conclusions regarding the developmental trajectory of the FTE at the current stage, and further tests are called for. Going beyond Aslan and Bäuml (2016), the current study employed a consistent test format across lists, unrelated word lists as principal stimuli, which have been widely used in previous FTE studies (Bäuml & Kliegl, 2013; Pashler et al., 2013; Szpunar et al., 2008), and a larger sample size to further determine the existence or absence of the FTE in young children.

Reset-of-encoding theory

The reset-of-encoding (ROE) theory has been recently proposed to account for the FTE (Pastötter et al., 2018). It assumed that memory load gradually increases with an increase in encoded lists. But testing of these encoded lists, as opposed to additional study periods, promotes contextual list segregation. Such enhanced segregation reduces memory load, induces a reset of the encoding process for the subsequent new lists, making the encoding of later new lists as effective as the encoding of earlier ones (Pastötter et al., 2011).

There is direct evidence illustrating that prior interim tests during learning facilitate subsequent encoding of new information. In the study of Pastötter et al. (2011), participants were instructed to study five lists of 20 items and did either a test or a restudy task between the study of the lists. The FTE was observed, with superior List 5 recall in the test condition than in the restudy condition. Moreover, alpha power (8–14 Hz) increases from List 1 to List 5 encoding in the restudy condition but no such increase in the test condition. Considering increases of alpha power reflect an impoverished encoding process caused by increased memory load and inattention (Sederberg et al., 2006, Pastötter et al., 2008; Pastötter et al., 2011), this result indicates that interim testing resets the encoding process and provides greater memory capacity for storage of subsequent list (Pastötter et al., 2011).

Apart from this neurocognitive work, a recent behavioural work also supports the ROE view. Specifically, Pastötter et al. (2018) asked participants to study three lists of words in test and restudy conditions. Similarly, participants in the test condition were tested after studying each list, whereas in the restudy condition they restudied Lists 1 and 2 and were tested on List 3. By using a serial position analysis, the results demonstrated a FTE that changed with items' serial list position, that is, a larger recall enhancement was found in early List 3 items (i.e., Items 1–4) compared to the middle (i.e., Items 5–8) and late (i.e., Items 9–12) items. This study verified the ROE theory on an item-level basis.

However, the ROE process was only evident in the FTE in adults, and it remains unknown whether this process is responsible for the FTE in children. Therefore, the current study aims to fill this gap by investigating the ROE process in the FTE in children. Since this ROE is induced by list segregation processes that do not include strategic decisions, even younger children should benefit from it.

Overview of the current study

To contribute to this growing literature, we firstly investigated whether younger children, older children, and adults (as in Aslan & Bäuml, 2016) can benefit from the FTE using a matched test format (i.e., free recall) across lists. Secondly, to unravel the mechanisms underlying the effect, we conducted a serial position analysis (Pastötter et al., 2018) and evaluated whether the ROE process contributes to the FTE in children and adults.

Method

Participants

We conducted a power analysis to establish the required sample size using G*Power (Faul et al., 2007). With a power set at 0.85, a two-tailed value of $\alpha = 0.05$, and an effect size of Cohen's $d = 0.80$, the power analysis showed that we needed 30 participants per condition to obtain a significant FTE. Sixty-four younger children (M = 6.38 years old, SD = 0.49; 29 female), 64 older children (M = 8.35 years old, SD = 0.42; 30 female) and 64 adults (M = 20.75 years old, SD = 1.63; 32 female) took part in the experiment.

Data from 7 participants were excluded because 2 younger children did not follow the instructions, one older child in the restudy condition was an outlier whose correct List 4 recall was more than three standard deviations above the mean recall, and four adults' data were unsaved due to a programming error. The final sample included data from 185 participants: 62 younger children (M = 6.37 years old, SD = 0.49; 27 female), 63 older children (M = 8.34 years old, SD = 0.48; 30 female), and 60 adults (M = 20.6 years old, SD = 1.48; 30 female), with 31 younger children, 32 older children and 30 adults in the test condition, and 31 younger children, 31 older children and 30 adults in the restudy condition.

All children were native Chinese speakers recruited from an elementary school in Jiangxi province of China, and participated voluntarily. The adults were students recruited from Beijing Normal University. They had a normal or corrected-to-normal vision and were without neurological or psychiatric diseases (based on self-reports for adults and caregivers' reports for younger and older children). The children knew they participated in an experiment, and informed consent was obtained from either participant themselves (for adults) or their caregivers (for younger and older children). The study was approved by the ethics committee of Beijing Normal University. Adult participants received 15 RMB and children received a set of stationery as compensation.

Materials

Four study lists were constructed, each consisting of six unrelated concrete Chinese nouns (see the Table A1) drawn from word norms for children developed by Liu et al. (2011).

Experiment design and procedure

The experiment employed a 3 (age: younger children *vs*. older children *vs*. adults) × 2 (interim task: test *vs*. restudy) between-subjects design. Participants were individually tested in a sound-proofed room.

Participants were initially informed that the experiment was designed to test their memory and mathematical abilities. Specifically, they would learn four lists of words, and they were encouraged to remember as many words as they could, in preparation for a final cumulative recall test in which all words from all lists would be tested. They were also informed that, after studying each list and solving maths problems for 30 sec, the computer programme would randomly decide whether to give them a short test or offer them a restudy opportunity.

Following the initial instructions, the study words in each list were presented in a random order, 5 sec each, and with a 1-sec inter-stimulus interval. After studying each list, participants solved as many maths problems as they could in 30 sec. Immediately following, participants in the test condition took a free-recall test (i.e., orally recall as many words as they could from the just-studied list in any order they wished) within 30 sec. While those in the restudy condition restudied the 6 words in a new random order on each of Lists 1–3, but took a free-recall test on List 4. Thus, the procedure for the target List 4 was same between the two conditions, but it was different for Lists 1–3. Participants' verbal responses were automatically recorded by the computer programme. The cumulative test was not implemented to keep the experiment short and to reduce task demands on children.

Results

List 4 interim test results

We examined whether the FTE exists in younger children, older children and adults. The corresponding descriptive statistics were depicted in Figure 1a and Table 1. A 3 (age: younger children *vs*. older children *vs*. young adults) × 2 (interim task: test *vs*. restudy) between-subjects analysis of variance (ANOVA) revealed a main effect of age on List 4 interim test recall, $F(2, 179) = 161.56$, $p < .001$, $\eta_p^2 = .64$. Correct recall increased with age (younger children: $M = 0.27$, $SD = 0.17$; older children: $M = 0.31$, $SD = 0.26$; adults: $M = 0.80$, $SD = 0.19$). Bonferroni corrected pairwise comparisons showed numerical or significant differences between

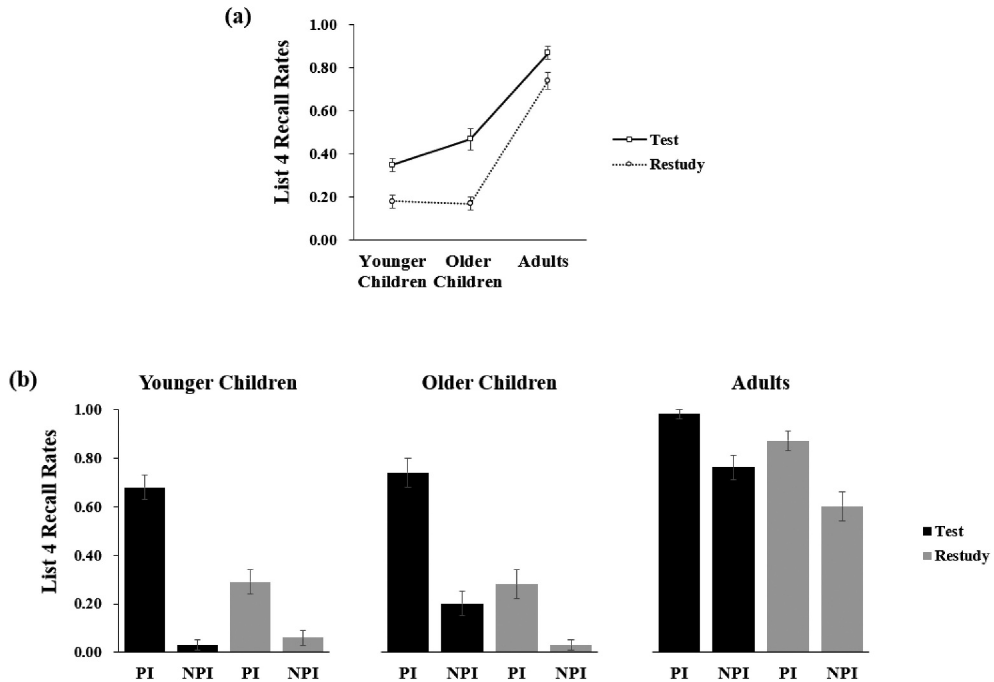

Figure 1. (a) Recall rates in the List 4 interim test as a function of age (younger children *vs*. older children *vs*. adults) and interim task (test *vs*. restudy). (b) Recall rates in the List 4 interim test as a function of interim task (test *vs*. restudy) and the items' serial list position (PI: primacy items 1 to 3, NPI: non-primacy items 4 to 6). Error bars represent ± 1 standard error.

Table 1. Proportion (*SD*) of Lists 1–4 interim test recall.

List	Interim task	Younger Children	Older Children	Adults
List 1	Test	0.45 (0.22)	0.52 (0.26)	0.84 (0.19)
List 2	Test	0.40 (0.20)	0.42 (0.23)	0.88 (0.12)
List 3	Test	0.28 (0.21)	0.44 (0.23)	0.84 (0.13)
List 4	Test	0.35 (0.15)	0.47 (0.25)	0.87 (0.16)
	Restudy	0.18 (0.14)	0.16 (0.15)	0.73 (0.20)

groups: younger children *vs*. older children: $p = .47$; $d = 0.26$; older children *vs*. adults: $p < .001$, $d = 2.67$; younger children *vs*. adults: $p < .001$; $d = 2.93$.

The main effect of interim task was significant, $F(1, 179) = 60.68$, $p < .001$, $\eta_p^2 = .25$, with higher recall in the test ($M = 0.56$, $SD = 0.29$) than in the restudy condition ($M = 0.35$, $SD = 0.31$), reflecting the FTE. Importantly, there was a significant interaction between the two factors, $F(2,179) = 4.12$, $p = .018$, $\eta_p^2 = .04$. A 2 (age: younger children *vs*. older children) × 2 (interim task: test *vs*. restudy) between-subjects analysis of variance (ANOVA) revealed a significant interaction between the two factors, $F(1, 121) = 4.39$, $p = .038$, $\eta_p^2 = .04$. Older children ($p < .001$, $d = 1.71$) had a larger FTE than younger children ($p < .001$, $d = 0.98$). A 2 (age: older children *vs*. adults) × 2 (interim task: test vs. restudy) between-subjects analysis of variance (ANOVA) revealed a significant interaction between the two factors, $F(1, 119) = 6.36$, $p = .013$, $\eta_p^2 = .05$. Older children had a larger FTE than adults ($p = .01$, $d = 0.67$). This result may be due to the fact that this task was too easy for adults, producing a celling effect driven by the higher levels of correct recall in the free-recall test in both the test and restudy conditions.

Interim test recall across lists

This section focuses on interim test recall evolvement 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 below analyses. Recall performance across lists for participants in the test condition was analysed by a 4 (Lists 1–4) × 3 (younger children *vs*. older children *vs*. adults) mixed ANOVA. The corresponding descriptive statistics were depicted in Table 1. Recall performance across lists differed significantly among the three groups, $F(2, 90) = 129.66$, $p < 0.001$, $\eta_p^2 = .74$. Bonferroni corrected post-hoc tests showed that recall performance increased with age: younger children *vs*. older children: $p = .016$, $d = 0.76$; older children *vs*. adults: $p < 0.001$, $d = 3.30$; younger children *vs*. adults: $p < 0.001$, $d = 4.10$.

Recall performance across lists differed significantly among the four lists, $F(3, 270) = 3.07$, $p = .028$, $\eta_p^2 = .03$. Bonferroni corrected post-hoc tests showed that there was a significant difference between List 1 and List 3: $p = .03$, $d = 0.42$, and there was no other significant difference between other lists, with *p*s ranging from .51 to .99 and *d*s ranging from 0.01 to 0.26.

There were no significant interaction between list and age ($F(6, 270) = 1.52$, $p = .17$, $\eta_p^2 = .03$). A repeated measures ANOVA revealed recall across lists remained stable in the test condition for older children: $F(3, 93) = 1.14$, $p = .34$, $\eta_p^2 = .04$, and adults: $F(3, 87) = 0.45$, $p = .72$, $\eta_p^2 = .02$, but test performance significantly varied across lists for younger children: $F(3, 90) = 4.23$, $p = .008$, $\eta_p^2 = .12$. Specifically, interim test recall linearly decreased across lists for younger children, $F(1, 30) = 8.30$, $p = .007$, $\eta_p^2 = .22$, suggesting that interim testing cannot fully prevent the decrease of learning efficiency across lists for younger children.

List 4 recall as a function of items' serial position

We examined whether participants in the three age groups recalled more List 4 primacy items than for the List 4 non-primacy items in the test condition compared to the restudy condition. The corresponding descriptive statistics were depicted in Figure 1b and Table 2. A 2 (interim task: test *vs*. restudy) × 2 (serial position: primacy items *vs*. non-primacy items) × 3 (age: younger children *vs*. older children *vs*. adults) mixed ANOVA revealed the main effects of interim task, $F(1, 179) = 60.68$, $p < .001$, $\eta_p^2 = .25$, serial position, $F(1, 179) = 222.73$, $p < .001$, $\eta_p^2 = .55$, and age, $F(2, 179) = 161.56$, $p < .001$, $\eta_p^2 = .64$.

Of critical interest, the interaction among condition, serial position, and age was also significant, $F(2, 179) = 8.22$, $p < .001$, $\eta_p^2 = .08$. Specifically, compared to recall in the restudy condition, primacy items benefited more from interim testing than non-primacy items in younger children (for primacy items: $p < .001$, $d = 1.47$; for non-primacy items: $p = .57$, $d = 0.14$) and older children (for primacy items: $p < .001$, $d = 1.76$; for non-

Table 2. Proportion (*SD*) of primacy and non-primacy items in List 4 interim test.

Interim task	Serial Position	Younger Children	Older Children	Adults
Test	PI	0.68 (0.27);	0.74 (0.32)	0.98 (0.08)
	NPI	0.03 (0.10);	0.20 (0.28)	0.76 (0.29)
Restudy	PI	0.29 (0.28);	0.28 (0.32)	0.87 (0.21)
	NPI	0.06 (0.16);	0.03 (0.10)	0.60 (0.31)

Note: PI: primacy items 1 to 3, NPI: non-primacy items 4 to 6.

primacy items: p = .004, d = 0.75), but not in adults (for primacy items: p = .10, d = 0.42; for non-primacy items: p = .008, d = 0.69). It may due to ceiling for primacy items in adults. Overall, these results expand on Pastötter et al.'s (2018) serial position findings, and can be taken as evidence for the ROE explanation for the FTEs of both younger and older children.

General discussion

Several important findings emerged from the current study. First and foremost, the FTE was robust in younger children, older children, and adults, as evidenced by the analyses of List 4 interim test recall. It is a nontrivial finding in light of the potential educational applications of the FTE. Second, serial list position significantly modulated the FTE in both younger children and older children, with a larger enhancement effect for the early List 4 items at list primacy positions compared to end List 4 items at non-primacy positions. Together, the present study tests and provides direct evidence for the ROE theory in younger and older children, indicating that the FTE in children may be driven by the ROE process. Below, we consider the theoretical implications of these findings.

Interim testing boosts new learning for younger children, older children, and adults

The finding that the interim testing of previously studied material enhanced new learning performance in all age groups represents a novel contribution to the literature: the FTE is efficient in the early elementary school years. Previous findings that showed the FTEs exist in young adults with varying levels of working memory capacity (Pastötter & Frings, 2019) and test anxiety (Yang et al., 2020), older adults (Pastötter & Bäuml, 2019), and clinical populations (Pastötter et al., 2013), together with the present results suggest that the FTE is broadly present in different populations. It should be noted that compared to older children and adults whose correct recall was stable across lists, younger children had a linear downward trend in recall from Lists 1–4. Thus, although the interim test promoted List 4 recall for all age groups, it cannot fully prevent the decrease of learning efficiency across lists for younger children.

These results are inconsistent with Aslan and Bäuml (2016) study, which showed that the FTE generalizes to older children and adults, but not to younger children. One procedural difference may contribute to this divergence. In Aslan and Bäuml (2016)'s study, the target list test (free-recall) mismatched with and is more difficult (accordingly more effortful retrieval required) than prior nontarget tests (cued-recall). Participants who received prior cued-recall tests before attending to a target free-recall test may be underprepared for the more difficult free recall test, due to the perceived ease of the previous tests. Compared to older children and adults, younger children might be more susceptible to this increased test difficulty and required retrieval effort from prior cued-recall tests to the critical free-recall test, resulting in that their FTE was not observed in Aslan and Bäuml (2016). However, when the target list test matched with prior nontarget lists test (e.g., both the tests are free-recall tests in the current study), participants, especially younger children, might be adequately prepared for the type of the target test, which in turn resulted in equivalent criterial test performance (Davis, 2018). This may lead to the result that their FTE were observed in the current study.

Other potential explanations for the differences in results between the current study and Aslan and Bäuml (2016)'s study arise from both retrieval strategy and retrieval effort accounts (Cho et al., 2017) of the FTE. According to the retrieval strategy account (Chan et al., 2020; Chan, Manley et al., 2018; Cho et al., 2017), younger children in the current study may gradually develop more effective retrieval strategies across the interim tests, such as temporal clustering strategy (Yang et al., 2021), and these strategies facilitate recall of subsequent target list. But this retrieval strategy mechanism may not contribute to Aslan and Bäuml (2016)'s study because different formats of retrieval require different retrieval strategies. According to the retrieval effort account (Cho et al., 2017), retrieval failures in prior interim tests may motivate young children to commit more effort to retrieve the words in the target list test in the current study. As the cued recall test is easier than the free recall test, younger children may experience fewer retrieval failures in prior nontarget tests in Aslan and Bäuml (2016)'s study than in the current study, resulting in less retrieval effort committed to retrieve the words in target list test.

In sum, when keeping test formats and test difficulty consistent across nontarget and target tests, the current findings contradict Aslan and Bäuml (2016) conclusion that the FTE is a relatively late-maturing phenomenon that develops over middle childhood (Aslan & Bäuml, 2016). Rather, the FTE develops quite early in life. Future studies should manipulate the criterial test type (cued recall or free recall) and task type (e.g., cued-recall, free recall, or no-test) to further determine whether prior cud-recall testing is not as potent as free-recall to produce the FTE in younger children.

Testing potentiates new learning that varied with items' serial list position

The finding that serial list position significantly modulated the FTE in both younger children and older children well parallels and extends earlier research on the ROE account of the FTE. Prior research (Pastötter et al., 2018) has found that in adults, interim testing induced a selective recall enhancement for the primacy items when compared to the restudy condition. We further showed that it occurred in both younger children and older children, that is, compared to recall in the restudy condition, primacy items benefited more from interim testing than non-primacy items. It suggests that the FTE in children was driven by the ROE process. Because the ROE is caused by contextual list segregation which does not involve strategic decisions, not only older children but also younger children benefit from it and present the FTE.

Specifically, interim testing rather than restudy promotes contextual list segregation. Such enhanced segregation abolishes or at least reduces memory load, induces encoding reset and makes subsequent new learning as effective as earlier learning. Moreover, as new learning goes on, memory load and inattention gradually increase, attenuating these benefits at later phases of new encoding and resulting in a smaller and even null effect for non-primacy items in older children and younger children, respectively. The results are consistent with studies which indicate that

contextual change influences young children's memory performance (e.g., Bartlett et al., 1982; Hala et al., 2013). Furthermore, the ROE account is not restricted to the FTE in younger children but has been applied to directed forgetting tasks (e.g., Hupbach et al., 2018), so the ROE may similarly affect retrieval in young children. In addition, adults' s List 4 recall for primacy items were nearly at the ceiling, making it impossible to observe such enhancement effect in the current study.

Because the ROE is caused by contextual list segregation which does not involve strategic decisions, preschoolers may also benefit from it and present the FTE. Future work should attempt to replicate the above findings and to investigate whether the FTE exists in preschoolers by using age-appropriate study materials and simplified experiment instructions. Moreover, others accounts of the FTE, such as retrieval strategy account and retrieval effort account as we mentioned before may also play roles in the FTE in children. Future studies could take into account the possible contributions and interactions of the multiple mechanisms of the FTE in children.

Conclusion

To summarize, immediate testing of studied lists enhanced correct recall of new learning in three age groups. It suggests that the FTE develops quite early in life and it generalizes to younger children, older children, and adults. Critically, serial list position significantly modulated the FTE in both younger children and older children, early List 4 items at list primacy positions benefited more from interim testing compared to late List 4 items at non-primacy positions, suggesting the ROE process contributes to the FTE in children. Educators can build on these results by strategically implementing interim tests across the course of a study session to promote young learners' learning of new information.

Acknowledgments

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

Disclosure statement

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

Funding

This work was supported by the major project grants from the National Social Science Foundation of China [14ZDB160].

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), 6. https://doi.org/10.1111/desc.12340

Bartlett, J. C., Burleson, G., & Santrock, J. W. (1982). Emotional mood and memory in young children. *Journal Of Experimental Child Psychology*, *34*(1), 59–76. https://doi.org/10.1016/0022-0965(82)90031-5

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

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*(2020), 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*(2018), 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

Davis, S. D. (2018). “Can multiple-choice testing potentiate new learning for text passages? A meta-cognitive approach to understanding the forward testing effect” (2018). *Graduate Theses and Dissertations*, 16566. https://lib.dr.iastate.edu/etd/16566

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

Hala, S., Brown, A. B., McKay, L., & San Juan, V. (2013). Two-and-a-half-year-olds’ memory for sources of actions: Contextual support facilitates recall. *Journal Of Cognition And Development*, *14*(2), 343–358. https://doi.org/10.1080/15248372.2012.664594

Hupbach, A., Weinberg, J. L., & Shiebler, V. L. (2018). Forget-me, forget-me-not: Evidence for directed forgetting in preschoolers. *Cognitive Development*, *45*(2018), 24–30. https://doi.org/10.1016/j.cogdev.2017.11.002

Liu, Y., Hao, M., Li, P., Shu, H., & Bishop, D. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS One*, *6*(1), e16505. https://doi.org/10.1371/journal.pone.0016505

Otgaar, H., Chan, J. C. K., Calado, B., & La Rooy, D. (2019). Immediate interviewing increases children’s suggestibility in the short term, but not in the long term. *Legal and Criminological Psychology*, *24*(1), 24–40. https://doi.org/10.1111/lcrp.12137

Pashler, H., Kang, S. H., & Mozer, M. C. (2013). Reviewing erroneous information facilitates memory updating. *Cognition*, *128*(3), 424–430. https://doi.org/10.1016/j.cognition.2013.05.002

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

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/10.3389/fpsyg.2018.01197

Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of Cognition*, *2*(1), 1–15. https://doi.org/10.5334/joc.82

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(2), 287–297. https://doi.org/10.1037/a0021801

Pastötter, B., Weber, J., & Bäuml, K.-H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, *27*(2), 280–285. https://doi.org/10.1037/a0031797

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy e!ect in episodicmemory. *Neuroimage*, *32*(3), 1422–1431. https://doi.org/10.1016/j.neuroimage.2006.04.223

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*, *110*(16),6313–6317. https://doi.org/10.1073/pnas.1221764110

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

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), 1-9. https://doi.org/10.1038/s41539-018-0024-y

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*, *26*(4), 724–738. https://doi.org/10.1037/xap0000278

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2021). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001021

Table A1. Twenty-four items used as stimuli in Chinese and English translation.

蛋糕Cake	屋顶Roof	地图Map
树叶Leaf	话筒Microphone	火车Train
茶杯Teacup	耳朵Ear	草莓Strawberry
月亮Moon	花生Peanut	闹钟Alarm
公路Road	手表Watch	窗帘Curtain
剪刀Scissor	台灯Lamp	萝卜Carrot
玉米Corn	瓶子Bottle	雨伞Umbrella
篮球Basketball	毛衣Sweater	小鸟Bird