Developmental Science

Developmental Science 19:6 (2016), pp 992–998 DOI: 10.1111/desc.12340

PAPER

Testing enhances subsequent learning in older but not in younger elementary school children

Alp Aslan[1,2] and Karl-Heinz T. Bäuml[1]

1. *Department of Experimental Psychology, Regensburg University, Germany*
2. *Department of Psychology, Martin Luther University Halle-Wittenberg, Germany*

Abstract

In adults, testing can enhance subsequent learning by reducing interference from the tested information. Here, we examined this forward effect of testing in children. Younger and older elementary school children and adult controls studied four lists of items in anticipation of a final cumulative recall test. Following presentation of each of the first three lists, participants were immediately tested on the respective list, or the list was re-presented for additional study. Results revealed that, compared to additional study, immediate testing of Lists 1–3 enhanced memory for the subsequently studied List 4 in adults and older elementary school children, but not in younger elementary school children. The findings indicate that the forward effect of testing is a relatively late-maturing phenomenon that develops over middle childhood and is still inefficient in the early elementary school years. Together with the results of other recent studies, these findings point to a more general problem in young children in combating interference.

Research highlights

- In older elementary school children, testing enhances the learning of new information by reducing interference from the tested memories.
- In younger elementary school children, this beneficial forward effect of testing is not yet present.
- Results suggest that the forward effect of testing is a relatively late-maturing phenomenon that develops over middle childhood.
- Together with other work, results point to a more general problem in young children in combating interference.

Introduction

Recent research has highlighted the critical importance of testing for improving learning and retention. Indeed, numerous studies have shown that taking a memory test on previously studied information can enhance long-term retention of that information more than repeated study of the information can do (for a review, see Roediger & Butler, 2011). This beneficial effect of testing, hereinafter referred to as the *backward effect* of testing (Pastötter & Bäuml, 2014), demonstrates what Roediger and Karpicke (2006) called a direct benefit of testing. It has been found in a number of settings and using a wide range of materials (e.g. Butler & Roediger, 2007; Hogan & Kintsch, 1971; Karpicke & Roediger, 2008; McDaniel, Anderson, Derbish & Morrissette, 2007; Roediger & Karpicke, 2006; Wheeler & Roediger, 1992).

However, there are also indirect benefits of testing. A particularly striking indirect benefit, hereinafter referred to as the *forward effect* of testing (Pastötter & Bäuml, 2014), is the finding that recall testing of previously studied information can increase long-term retention of subsequently studied new information. In the original study on this effect, Szpunar, McDermott and Roediger (2008) let each of three groups of participants learn five lists of items in anticipation of a final cumulative recall test. Participant groups differed in the activity that followed presentation of each of the first four (nontarget)

Address for correspondence: Alp Aslan, Department of Psychology, Martin Luther University Halle-Wittenberg, Emil-Abderhalden-Strasse 26-27, 06108 Halle (Saale), Germany; e-mail: alp.aslan@psych.uni-halle.de

lists, with one (control) group doing a filler task after Lists 1–4, one group restudying Lists 1–4, and one group being immediately tested on Lists 1–4. Results revealed that immediate testing of Lists 1–4 improved later recall of the subsequently studied (target) List 5, with the tested group recalling about twice as many List 5 items as the two other groups. In addition, immediate testing of Lists 1–4 reduced number of intrusions from these lists during recall of List 5, relative to both the restudy and the control condition, suggesting that testing during study insulates against the build-up of proactive interference. Subsequent work replicated these findings (e.g. Pastötter, Schicker, Niedernhuber & Bäuml, 2011) and extended them to a variety of other study materials (e.g. Pastötter, Weber & Bäuml, 2013; Szpunar, Khan & Schacter, 2013; Weinstein, McDermott & Szpunar, 2011; Wissman, Rawson & Pyc, 2011; for a recent review, see Pastötter & Bäuml, 2014).

Theoretical accounts of the forward effect of testing differ in whether they emphasize encoding or retrieval processes. Retrieval-based explanations assume that testing of previously studied nontarget materials enhances segregation between nontarget and target lists more than restudy does (Szpunar *et al.*, 2008). Such enhanced segregation may then reduce mental search set size when target items are recalled and allow more focused memory search of the target list items (Bäuml & Kliegl, 2013; Lehman, Smith & Karpicke, 2014). In contrast, encoding-based explanations assume that testing of previously studied nontarget materials already improves the encoding of the subsequently studied target material. This may occur because testing induces a reset of the encoding process, making the encoding of the later lists as effective as the encoding of the earlier lists (Pastötter *et al.*, 2011), or because testing induces a change in encoding strategy, enhancing elaborative encoding for the later lists compared to the earlier lists (Wissman *et al.*, 2011). Together, the results suggest that both encoding and retrieval processes may be involved when testing during study insulates against the build-up of proactive interference.

Although the forward effect of testing has proved to be a very general and robust finding in young adults, little is yet known about whether the effect generalizes to populations other than undergraduate students (for an exception, see Pastötter *et al.*, 2013). In particular, despite its obvious relevance for educational settings, to date no study has yet examined whether the effect is also present in school-aged children. This is surprising given that children have to learn and retain vast amounts of information from the very first days of their school lives, and practitioners would appreciate having instructional tools and techniques available to support children in this endeavor. Specifically, both teachers and parents would like to know whether, in elongated study sessions in which several sets of information have to be learned in succession, children's learning and retention of new, yet-to-be-studied topics could be facilitated by providing an interim test for other, earlier-studied topics.

That children *can* benefit from tests has been demonstrated in previous research on the development of the backward effect of testing. Indeed, although the backward effect of testing has mostly been examined in adult samples, recent developmental work has established the effect also in younger participants, including middle school children (Carpenter, Pashler & Cepeda, 2009; McDaniel, Agarwal, Huelser, McDermott & Roediger, 2011; Rohrer, Taylor & Sholar, 2010), elementary school children (Bouwmeester & Verkoeijen, 2011; Lipowski, Pyc, Dunlosky & Rawson, 2014), and even preschool children as young as 3 to 5 years of age (Fritz, Morris, Nolan & Singleton, 2007). These findings indicate that the mechanisms underlying the backward effect of testing develop quite early in life and are intact at a relatively young age. However, whether the developmental trajectory of the one (backward) effect of testing simply generalizes to the other (forward) effect of testing, so that not only adults but also young children show the benefits of testing on the learning of subsequent information, is unclear.

Indeed, indirect evidence from a recent series of experiments raises the possibility that the developmental trajectories of the two testing effects differ and the forward effect of testing develops later than the backward effect. These experiments repeatedly found that young children can have great difficulty in dealing efficiently with proactive interference. Employing the directed forgetting task, for instance, it was found that instructing participants to forget a previously studied item list before learning a new list of items reduced proactive interference from the first list and enhanced recall of the second list items in young adults and fourth graders, but not in first graders and kindergartners (Aslan, Staudigl, Samenieh & Bäuml, 2010b; Harnishfeger & Pope, 1996). Similar results were reported when employing an environmental context change, instead of a forget cue, between the study of two lists (Aslan, Samenieh, Staudigl & Bäuml, 2010a). Finally, it was shown that inserting an imagination task between study of two lists, supposedly inducing a change in the subjects' mental context, improved later recall of the second list in young adults, but not in elementary school children or kindergartners (Aslan & Bäuml, 2008). Together, these results suggest that younger, and sometimes even older, children can have greater difficulty than adults in combating proactive interference from previously studied information. If this difficulty was not specific to directed forgetting or context change settings, but was more general in nature, then the forward effect of testing might also be diminished in school-aged children.

The goal of the present study was to examine whether school-aged children show the forward effect of testing. To achieve this goal, we let younger and older elementary school children and adult controls study four lists of items in anticipation of a final cumulative recall test. After each of the first three (nontarget) lists, participants either restudied the just-presented list, or were given a cued-recall test on the list. The critical variable was performance on (target) List 4, which was always assessed using free recall. In young adults, we expected to replicate previous work (e.g. Pastötter *et al.*, 2011; Szpunar *et al.*, 2008), and find testing of Lists 1–3 to enhance List 4 recall and reduce number of intrusions from Lists 1–3, relative to restudying. If previous findings on the development of the backward effect of testing (e.g. Fritz *et al.*, 2007; Lipowski *et al.*, 2014) generalized to the forward effect, these beneficial effects should also be found in the two children's groups. Alternatively, if children suffered from a more general deficit in combating proactive interference, that was not restricted to directed forgetting or context change settings (e.g. Aslan *et al.*, 2010a, 2010b), then the forward effect of testing might be reduced, or even be eliminated, in school-aged children.

Method

Participants

Thirty-two younger children (M = 6.7, SD = 0.5 years), 32 older children (M = 8.8, SD = 0.8 years), and 32 adults (M = 23.3, SD = 3.3 years) took part in the experiment. The children were recruited from two elementary schools near Regensburg and Munich, Germany, and participated on a voluntary basis; the adults were students of Regensburg University and received course credit. All participants were tested individually. [Correction added on 09 August 2016, after first online publication: The number of participants has been corrected from 48 to 32.]

Materials

Four study lists were constructed, each consisting of six unrelated concrete German nouns drawn from word norms for children (Hasselhorn, Jaspers & Hernando, 1990; Posnansky, 1978). All items began with a unique word stem.

Design and procedure

The experiment had a 3 × 2 between-subjects design with the factors of Age Group (younger children, older children, adults) and Condition (testing, restudy).

At the beginning of the experiment, participants were informed that they would be presented with several lists of items, which should be studied and kept in mind for a possible final cumulative recall test for all lists.[1] Following the initial instruction, the four lists were presented orally by the experimenter. The experimenter read out the six items of each list at a 5-sec rate in random order. Presentation of each list thus took approximately 30 sec, and was immediately followed by an (irrelevant) 30-sec trail-making task as a filler. The two learning conditions differed in the activity that followed the filler task after Lists 1–3. Participants in the restudy condition were orally re-presented the just-studied list's six items in a new random order for an additional study trial; participants in the testing condition were orally presented the words stems (2–4 letters, depending on word length) of the just-presented list's six items in random order, and were asked to complete the corresponding words within 5 sec each. Recall of the nontarget lists was cued with the items' word stems to increase performance on these tests and thus boost possible effects of testing on List 4 performance. The procedure for the critical List 4 was the same for the two learning conditions, but differed from the procedure for Lists 1–3 in that, after list presentation and the 30-sec filler task, all participants were given 30 sec to recall in any order they wished as many words as possible from List 4. Participants' verbal responses were noted by the experimenter. The whole experiment lasted approximately 15 min.

Results

Recall of Lists 1–3

As shown in Table 1, immediate word-stem-cued recall of Lists 1–3 was relatively high in all three age groups. A 3 × 3 analysis of variances (ANOVA) with the within-subjects factor of List (List 1, List 2, List 3) and the between-subjects factor of Age Group (younger children, older children, adults) revealed no main effect of List (List 1: 89.9%; List 2: 89.2%; List 3: 90.3%), $F(2, 90) = 0.12$, $MSE = 0.011$, $p = .888$, $\eta_p^2 = .003$, a

[1] A possible final cumulative test for all lists was announced to ensure that participants in the testing condition would continue to rehearse the nontarget Lists 1–3 after having been tested on the lists. Actually, however, the experiment ended with the free-recall test of List 4 to keep the experiment short and reduce task demands on children.

Table 1 *Mean percentages of recall (M), standard deviations (SD), and standard errors (SE) for Lists 1–3 in the testing condition as a function of Age Group (younger children, older children, adults)*

	List 1			List 2			List 3		
	M	*SD*	*SE*	*M*	*SD*	*SE*	*M*	*SD*	*SE*
Younger children	85.4	13.4	3.4	87.5	9.6	2.4	88.5	11.7	2.9
Older children	90.6	12.1	3.0	88.5	11.7	2.9	90.6	10.5	2.6
Adults	93.8	10.3	2.6	91.7	8.6	2.2	91.7	8.6	2.2

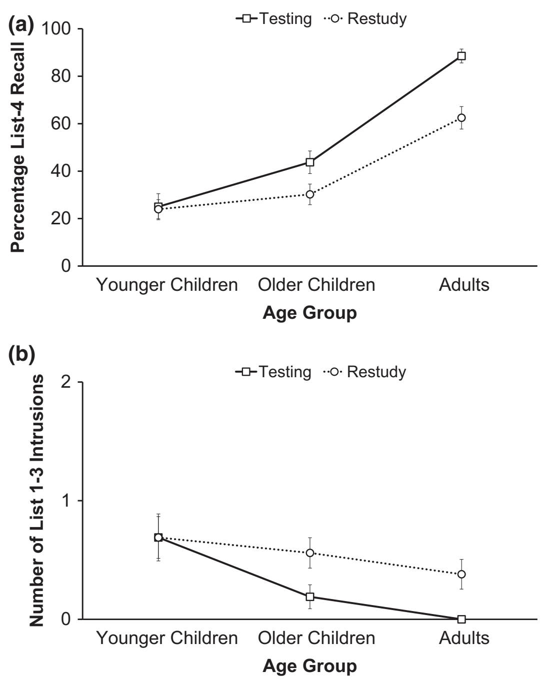

Figure 1 *(a) Mean percentage of List 4 recall as a function of Age Group (younger children, older children, adults) and Condition (testing, restudy). The error bars represent standard errors. (b) Mean number of prior list intrusions from Lists 1–3 during recall of List 4 as a function of Age Group (younger children, older children, adults) and Condition (testing, restudy). The error bars represent standard errors.*

marginally significant main effect of Age Group (younger children: 87.1%; older children: 89.9%; adults: 92.4%), $F(2, 45) = 2.57$, $MSE = 0.013$, $p = .088$, $\eta_p^2 = .103$, and no interaction between the two factors, $F(4, 90) = 0.32$, $MSE = 0.011$, $p = .862$, $\eta_p^2 = .014$.

Recall of List 4

Recall performance of List 4 is shown in Figure 1a (see also Table S1 in the Supporting Information, for details). A 3 × 2 ANOVA with the between-subjects factors of Age Group (younger children, older children, adults) and Condition (testing, restudy) revealed a significant main effect of Age Group, $F(2, 90) = 60.87$, $MSE = 0.037$, $p < .001$, $\eta_p^2 = .575$, reflecting higher overall recall with increasing age (younger children: 24.5%; older children: 37.0%; adults: 75.5%; all three pairwise comparisons: $p < .020$, $d > 0.65$), and a significant main effect of Condition, $F(1, 90) = 11.83$, $MSE = 0.037$, $p = .001$, $\eta_p^2 = .116$, reflecting higher overall recall in the testing (52.4%) than in the restudy condition (38.9%). Importantly, there was a significant interaction between the two factors, $F(2, 90) = 3.36$, $MSE = 0.037$, $p = .039$, $\eta_p^2 = .069$, indicating that the beneficial forward effect of testing varied with age. Indeed, subsequent analyses revealed that whereas the effect was quite pronounced in adults (26.0%; $t(30) = 3.63$, $SE = 0.072$, $p = .001$, $d = 1.28$), and still present in older children (13.5%; $t(30) = 2.09$, $SE = 0.065$, $p = .045$, $d = 0.74$), it was absent in younger children (1.0%; $t(30) = 0.15$, $SE = 0.068$, $p = .879$, $d = 0.05$).

Prior list intrusions

As shown in Figure 1b, there were relatively few prior-list intrusions from Lists 1–3 during recall of List 4. Still, a 3 × 2 ANOVA with the between-subjects factors of Age *Group* (younger children, older children, adults) and Condition (testing, restudy) revealed a significant main effect of Age Group, $F(2, 90) = 6.81$, $MSE = 0.300$, $p = .002$, $\eta_p^2 = .131$, reflecting the fact that younger children made more intrusion errors (0.69) than both older children (0.38), $p = .025$, $d = 0.49$, and adults (0.19), $p < .001$, $d = 0.84$, and a significant main effect of Condition, $F(1, 90) = 5.00$, $MSE = 0.300$, $p = .028$, $\eta_p^2 = .053$, reflecting lower overall intrusion rates in the testing (0.29) than in the restudy condition (0.54). Although the interaction between the two factors did not reach significance, $F(2, 90) = 1.25$, $MSE = 0.300$, $p = .291$, $\eta_p^2 = .027$, likely because intrusion rates in the testing condition were at or close to floor in adults (0.00) and older children (0.19), planned comparisons revealed that, relative to restudying, testing significantly reduced intrusion rates in adults ($t(30) = 3.00$, $SE = 0.125$, $p = .005$, $d = 1.07$) and older children ($t(30) = 2.30$, $SE = 0.163$, $p = .029$, $d = 0.80$), but not in younger children ($t(30) < 1$, $d = 0.00$).

Discussion

Previous work with adults has shown that testing can enhance subsequent learning by reducing proactive interference from previously studied information (e.g. Szpunar *et al*., 2008). We replicated this work by finding immediate testing of previously studied nontarget lists to both enhance adults' recall of a subsequently studied target list, and reduce adults' number of intrusions from the nontarget lists during target list recall. Going beyond the previous work, we found these effects to be numerically reduced, though still significant, in older elementary school children, and to be absent in younger elementary school children. These findings indicate that the forward effect of testing is a relatively late-maturing phenomenon that develops over middle childhood and is still inefficient in the early elementary school years.

Previous work with adults provided evidence that both encoding and retrieval processes can contribute to the forward effect of testing. For instance, employing response latency analysis, Bäuml and Kliegl (2013) found immediate testing of previously studied (nontarget) information to accelerate recall of subsequently studied (target) information, relative to restudy of the nontarget information. Because response latency is often positively related to the size of participants' memory search set (e.g. Wixted & Rohrer, 1994), the results suggest that testing induces more focused memory search and, during participants' retrieval of the target list, delimits the search set to the relevant target information (for similar results, see Lehman *et al*., 2014). Measuring participants' electroencephalogram during encoding of target and nontarget lists, Pastötter *et al*. (2011) found that immediate testing of the previously studied nontarget lists disrupts the increase in alpha power (10–14 Hz) that is typically observed with increasing amounts of study materials, thus leading to similar alpha levels for the target and nontarget lists. Because alpha power increase has been associated with impoverished encoding due to increased inattention and memory load (e.g. Klimesch, 2012; Palva & Palva, 2007), the finding indicates that testing can improve subsequent learning by keeping attention during encoding high and memory load low (for related results, see Pastötter, Bäuml & Hanslmayr, 2008). On the basis of these previous findings, the present results suggest that neither the ability to focus memory search during retrieval nor the ability to reset the encoding process after prior nontarget encoding is operative in young elementary school children. At least one of the two abilities seems to become functional over the elementary school years, enabling older children to show the forward effect of testing. The finding that the effect was numerically only half the size in older children compared with adults, however, suggests further maturation of these processes beyond the elementary school years.

Recent work with adults suggests that the forward effect of testing may also be affected by test expectancy. For instance, despite the instruction to anticipate a final cumulative recall test for all lists, the immediate testing of the nontarget lists in the testing condition may provide a 'sense of closure' to these lists. Relative to the restudy condition, such 'sense of closure' may reduce participants' expectation of a future test, leading to reduced processing of the nontarget lists and thus less proactive interference when recalling the target list items (Szpunar, McDermott & Roediger, 2007). Alternatively, participants in the restudy condition, who are never tested during study of the nontarget lists, may gradually cease expecting a test and thus reduce attentional processing when it comes to studying the target list. Because no such reduction in attentional processing should occur in the testing condition, in which subjects receive immediate tests after study of each single nontarget list, recall may be higher in the testing than in the restudy condition (Weinstein, Gilmore, Szpunar & McDermott, 2014). To the degree that such high-level reasoning about the likelihood of later tests is involved in the present situation, younger children might be expected to be poorer at this reasoning and fail to show the memorial benefits that can result from interim testing in older children and adults. Future work would be required to examine the possible role of test expectancy in the forward effect of testing in children directly.

Previous developmental research has repeatedly found that young children can have difficulty in combating interference from previously encoded information when new information has to be learned and recalled. Indeed, whereas such proactive interference is typically reduced in adults when a cue is provided to forget the previous information (e.g. Bjork, 1989) or an environmental or mental context change takes place before the new information is learned (Sahakyan & Kelley, 2002; Smith & Vela, 2001), young children did not show any reduction of proactive interference in these types of tasks (Aslan & Bäuml, 2008; Aslan *et al*., 2010a, 2010b; Harnishfeger & Pope, 1996). The present finding that the beneficial effect of interpolated testing on subsequent learning is numerically reduced in older elementary school children, and is even absent in younger elementary school children, fits with the results of these previous studies. Together the findings suggest that inability to combat proactive interference is a fairly general deficit in young children's cognition that is not restricted to specific paradigms but may generalize across a wide range of tasks.

The present finding of a relatively late maturation of the forward effect of testing, first of all, contrasts with previous work on the development of the backward effect of testing. In fact, this previous work found that the backward effect develops rather early in life, and is present in children as young as preschoolers (e.g. Fritz *et al.*, 2007). Importantly, although the exact processes underlying the backward effect of testing are still a matter of debate, theoretical accounts have often proposed mechanisms that differ from those proposed to underlie the forward effect, including enhanced elaboration (Carpenter & DeLosh, 2006; Pyc & Rawson, 2010), enhanced gist processing (Bouwmeester & Verkoeijen, 2011), or enhanced strengthening (Kornell, Bjork & Garcia, 2011) of the tested material. The present results thus are not necessarily in conflict with the previous findings. Rather, the developmental dissociation between the forward and the backward effects of testing suggests that the two effects of testing are at least partly mediated by different mechanisms with different developmental trajectories.

The goal of the present study was to examine the development of the forward effect of testing by using a relatively wide age range and the same study materials in all participant groups. In consequence, target recall was relatively low in the younger children and relatively high in adults. In addition, due to the use of unrelated study lists, intrusion rates from nontarget lists were relatively low in all three age groups, being at or close to floor in older children and adults, especially in the testing condition. Further work may therefore try to conceptually replicate and generalize the present findings using (i) procedures that ensure higher recall levels in children, like repeated study cycles for the single lists, and (ii) material that ensures higher intrusion rates in all age groups, like the use of semantically interrelated study lists (e.g. Szpunar *et al.*, 2008). With the goal of limiting the demands placed on the younger children, this study also did not include a final cumulative recall test. The omission of such a test leaves unclear whether the significant forward effect of testing found in the group of older children is transient or lasting. Unless highly cohesive prose material is used (Wissman & Rawson, 2015), the forward effect of testing in adults can be lasting and persist to delayed recall testing (e.g. Pastötter *et al.*, 2011; Szpunar *et al.*, 2008). Using both word lists and prose materials, future developmental work may therefore examine directly whether the persistency of the forward effect of testing holds similarly in adults and older children.

To conclude, this is the first study to examine the forward effect of testing in school-aged children, indicating that the memory of adults and older, but not younger, elementary school children can benefit from the immediate testing of previously studied information. The finding holds implications for both developmental research and educational practice. From a more theoretical research perspective, the finding may serve as a starting point for future work that, using both behavioral and neurocognitive methods, could clarify whether the two component processes supposed to underlie the forward effect of testing, i.e. the encoding-based and the retrieval-based mechanism, develop at the same rate or show different developmental trajectories. From a more practical educational perspective, the finding provides important hints and restrictions for the age-appropriate use of tests as instructional tools in basic education.

References

Aslan, A., & Bäuml, K.-H. (2008). Memorial consequences of imagination in children and adults. *Psychonomic Bulletin & Review*, **15**, 833–837.

Aslan, A., Samenieh, A., Staudigl, T., & Bäuml, K.-H.T. (2010a). Memorial consequences of environmental context change in children and adults. *Experimental Psychology*, **57**, 455–461.

Aslan, A., Staudigl, T., Samenieh, A., & Bäuml, K.-H.T. (2010b). Directed forgetting in young children: evidence for a production deficiency. *Psychonomic Bulletin & Review*, **17**, 784–789.

Bäuml, K.-H.T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, **68**, 39–53.

Bjork, R.A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H.L. Roediger & F.I.M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bouwmeester, S., & Verkoeijen, P.P.J.L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, **65**, 32–41.

Butler, A.C., & Roediger, H.L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, **19**, 514–527.

Carpenter, S.K., & DeLosh, E.L. (2006). Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, **34**, 268–276.

Carpenter, S.K., Pashler, H., & Cepeda, N.J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology*, **23**, 760–771.

Fritz, C.O., Morris, P.E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: an effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology*, **60**, 991–1004.

Harnishfeger, K.K., & Pope, R.S. (1996). Intending to forget: the development of cognitive inhibition in directed-forgetting. *Journal of Experimental Child Psychology*, **62**, 292–315.

Hasselhorn, M., Jaspers, A., & Hernando, M.-D. (1990). Typizitätsnormen zu zehn Kategorien von der Vorschule bis zur vierten Grundschulklasse. *Sprache & Kognition*, **9**, 92–108.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, **10**, 562–567.

Karpicke, J.D., & Roediger, H.L. (2008). The critical importance of retrieval for learning. *Science*, **319**, 966–968.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, **16**, 606–617.

Kornell, N., Bjork, R.A., & Garcia, M.A. (2011). Why tests appear to prevent forgetting: a distribution-based bifurcation model. *Journal of Memory and Language*, **65**, 85–97.

Lehman, M., Smith, M.A., & Karpicke, J.D. (2014). Toward an episodic context account of retrieval-based learning: dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**, 287–297.

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**, 994–1000.

McDaniel, M.A., Agarwal, P.K., Huelser, B.J., McDermott, K.B., & Roediger, H.L. (2011). Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *Journal of Educational Psychology*, **103**, 399–414.

McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, **19**, 494–513.

Palva, S., & Palva, J.M. (2007). New vistas for frequency band oscillations. *Trends in Neurosciences*, **30**, 150–158.

Pastötter, B., & Bäuml, K.-H.T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*, **5**, 286.

Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change: evidence for a reset of encoding processes. *NeuroImage*, **43**, 173–181.

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**, 287–297.

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

Posnansky, C.J. (1978). Category norms for verbal items in 25 categories for children in Grades 2–6. *Behavior Research Methods & Instrumentation*, **10**, 819–832.

Pyc, M.A., & Rawson, K.A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science*, **333**, 335.

Roediger, H.L., & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, **15**, 20–27.

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

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**, 233–239.

Sahakyan, L., & Kelley, C.M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **28**, 1064–1072.

Smith, S.M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin & Review*, **8**, 203–220.

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, USA*, **110**, 6313–6317.

Szpunar, K.K., McDermott, K.B., & Roediger, H.L. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition*, **35**, 1007–1013.

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

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**, 1039–1048.

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

Wheeler, M.A., & Roediger, H.L. (1992). Disparate effects of repeated testing: reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, **3**, 240–245.

Wissman, K.T., & Rawson, K.A. (2015). Grain size of recall practice for lengthy text material: fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 439–455.

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

Wixted, J.T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: an integrative review of the empirical literature. *Psychonomic Bulletin & Review*, **1**, 89–106.

Received: 11 February 2105
Accepted: 26 May 2015

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Mean percentages of recall (*M*), standard deviations (*SD*), and standard errors (*SE*) for List 4 as a function of *Age Group* (younger children, older children, adults) and *Condition* (testing, restudy).