

The interim test effect: Testing prior material can facilitate the learning of new material

Kathryn T. Wissman, Katherine A. Rawson & Mary A. Pyc

Psychonomic Bulletin & Review

ISSN 1069-9384

Psychon Bull Rev
DOI 10.3758/
s13423-011-0140-7

**Psychonomic
Bulletin &
Review**

VOLUME 17, NUMBER 3 ■ JUNE 2010

PB&R

EDITOR

Robert M. Nosofsky, *Indiana University, Bloomington*

ASSOCIATE EDITORS

Michael F. Brown, *Villanova University*

Thomas A. Busey, *Indiana University, Bloomington*

Richard J. Gerrig, *Stony Brook University*

Stephen D. Goldinger, *Arizona State University*

Ulrike Hahn, *Cardiff University*

Steven J. Luck, *University of California, Davis*

W. Trammell Neill, *University at Albany*

Jay Pratt, *University of Toronto*

Gillian Rhodes, *University of Western Australia*

Caren M. Rotello, *University of Massachusetts, Amherst*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1069-9384



Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

The interim test effect: Testing prior material can facilitate the learning of new material

Kathryn T. Wissman · Katherine A. Rawson ·
Mary A. Pyc

© Psychonomic Society, Inc. 2011

Abstract A wealth of prior research has shown that testing can improve subsequent learning of the initially tested material. In contrast, only one recent study has shown that an interim test over prior material can improve learning of subsequent new material (i.e., an *interim-test effect*). Five experiments replicated and extended this initial work by exploring the extent to which interim test effects generalize to complex text material. Participants were prompted to recall each section of an expository text before moving on to study the next section, or were only prompted to recall after the final section. In all experiments, recall of the final, target section was greater when prior sections had received interim tests versus no interim tests. **Experiment 3** established that the effect was due to interim testing in particular rather than to intervening activity in general. **Experiment 4** established that the effect was not due to test expectancy differences. In contrast to prior research, **Experiment 4** also provided evidence that the effect is not due to release from proactive interference. We discuss other possible mechanisms underlying interim-test effects with text, including shifting to more effective encoding strategies.

Keywords Testing effects · Text comprehension · Human memory and learning

K. T. Wissman · K. A. Rawson (✉) · M. A. Pyc
Department of Psychology, Kent State University,
P.O. Box 5190, Kent, OH 44242-0001, USA
e-mail: krawson1@kent.edu

Present Address:

M. A. Pyc
Department of Psychology, Washington University in St. Louis,
St. Louis, MO, USA

Previous research has established that testing during practice improves memory (Roediger & Butler, 2011). However, almost all previous research has shown item-specific benefits of testing. As illustrated at the top of Table 1, a common methodology used in the testing effect literature involves presenting target material (e.g., word pairs, expository text) for initial study, followed by either a practice test with restudy or restudy only for that same target material. Research has consistently shown that testing prior to restudy of a given item facilitates performance on a subsequent test of that item. This effect is due at least in part to the test facilitating more effective encoding of the information during subsequent restudy (*potentiating effects of testing*; e.g., Izawa, 1971; Karpicke, 2009).¹

In contrast to research showing that tests can potentiate subsequent learning of the same material, the present research addresses the intriguing question of whether an interim test over some initial material facilitates the learning of subsequent new material. As illustrated in Table 1, suppose that learners initially study Material A and then do or do not take an interim test over Material A prior to studying Material B. All learners are then tested over Material B. Does the interim test over Material A influence test performance for Material B?

In contrast to the sizeable literature on testing effects, only one recent study has directly examined what we refer to as the *interim-test effect* (i.e., the effect of taking an interim test over preceding material on the learning of subsequent new material). Szpunar, McDermott, and Roediger (2008) instructed participants to study five lists

¹ A considerable amount of previous research has also shown that a practice test can facilitate memory for the same material, even in the absence of subsequent restudy (e.g., Chan & McDermott, 2007; Darley & Murdock, 1971), although this paradigm is less comparable to the effects under consideration in the present research.

Table 1 Illustration of methods used in four related literatures examining the effects of testing on memory

Testing Effects					
Testing:	Study A	Test on A	Restudy A	Test on A	
No testing:	Study A		Restudy A	Test on A	
Interim-Test Effects					
Testing:	Study A	Test on A	Study B	Test on B	
No testing:	Study A		Study B	Test on B	
Retrieval-Induced Forgetting/Facilitation					
Testing:	Study A, B	Test subset of A		Test on A, B	
No testing:	Study A, B			Test on A, B	
Insulation Effect					
Testing:	Study A	Test on A	Study A'	Test on A, A'	
No testing:	Study A		Study A'	Test on A, A'	

of unrelated nouns. After each of the first four lists, participants completed 2 min of math problems or 1 min of math followed by a test over the prior list. After studying List 5, all participants recalled List 5. List 5 recall was significantly greater when prior lists had received interim tests versus no interim tests (39% vs. 19%). In follow-up experiments using interrelated word lists, Szpunar et al. found an even stronger interim-test effect (54% vs. 24%). Interestingly, intrusions of prior list items during List 5 recall were greater in the no-interim-test group than in the interim-test group (21% vs. 2%), suggesting that interim tests reduce proactive interference from earlier-learned lists.

Given that only one study has explored interim-test effects, the goal of the present research was to replicate and extend this initial work by exploring the extent to which these effects generalize to more complex text material. Much like Szpunar et al.'s (2008) interrelated word lists, sections within a text typically contain conceptually related content. This parallel suggests that the same pattern of facilitated recall might emerge when using text material, to the extent that interim tests over preceding sections reduce proactive interference for a subsequent target section, as for Szpunar et al.'s word lists. However, an important difference arises between interrelated word lists and interrelated text sections: In contrast to word lists, understanding the content in one text section often depends on integration with information contained in previous sections. Thus, interim tests may disrupt processing of a later text section by interfering with the activation of related information from preceding sections, which might compromise learning for the target section.

Although no prior research has examined interim-test effects for text material, recent findings from the related literature on retrieval-induced forgetting are indirectly relevant. Chan (2009, 2010; Chan, McDermott & Roediger 2006) adapted the standard method used to demonstrate retrieval-induced forgetting for word lists (see Table 1) to create an analogue using text material. Chan (2009)

presented learners with two articles for initial study, followed by retrieval practice for a subset of facts contained in one of the articles. Of interest here, the final test included questions about facts related to the ones that had been tested during practice. Final test performance for these related facts was facilitated by prior testing of their companion facts, relative to untested items in the second article (although retrieval-induced facilitation only occurred when the text material afforded a high degree of integration). However, as shown in Table 1, this procedure differs from the interim-test method, given that related material was studied before versus after recall of the tested information.

Findings from the literature on insulation effects may also be indirectly relevant. Using the A–B/A–C paradigm, Tulving and Watkins (1974) showed that a test of A–B items prior to study of A–C pairs improved subsequent recall of A–C items (i.e., the *insulation effect*). In a recent instantiation of this paradigm using more complex material (summarized in Table 1), Chan, Thomas, and Bulevich (2009, Exp. 2) had participants watch an episode of a television program depicting a plane hijacking, followed by either a cued recall test over the content or a filler task. All participants then listened to a short audio narrative of the events in the video that included several misleading items (e.g., stating that the terrorist knocked out the flight attendant with chloroform when the video actually portrayed a hypodermic injection). On the final cued recall test, participants were told to respond with any relevant information they could remember, regardless of the source. Recall of misleading information was greater for participants tested over the video versus those not tested. However, this design differs from the interim-test method, given that it uses modified versions of the same material versus entirely new information after the interim test.

In sum, although results from these other literatures are suggestive, no prior research has evaluated interim-test effects using complex material. Accordingly, the goal of the

present study was to explore the extent to which interim-test effects obtain with complex text material. In [Experiments 1A and 1B](#), expository texts were divided into three sections. Participants in the interim-test group were prompted to recall after reading each section. Participants in the no-interim-test group were not prompted to recall until after Section 3. [Experiments 2–4](#) explored alternative interpretations of the interim-test effects observed in [Experiments 1A and 1B](#).

Experiments 1A and 1B

Szpunar et al. (2008) showed that an interim-test effect could be obtained, regardless of the degree of relatedness between word lists. [Experiments 1A and 1B](#) involved texts that either were or were not related, in order to establish that interim-test effects with more complex materials also do not depend on degree of interrelatedness.

Method

Participants and design Undergraduates from Kent State University who participated for course credit were randomly assigned to one of two groups (interim test or no interim test; $ns = 64$ vs. 65 in Exp. 1A, $ns = 21$ vs. 19 in Exp. 1B).

Materials and procedure [Experiment 1A](#) included an expository text concerning forms of government intervention in the U.S. labor market (779 words, 12.0 Flesch grade level). The text was divided into three sections, each including a subtopic header (Benefit Mandates, Labor Laws, and Job Training Programs). The sections each described one form of government intervention, but otherwise were not directly related to one another and had no overlapping information. [Experiment 1B](#) included an expository text on capturing and storing atmospheric greenhouse gases (1,062 words, 12.0 Flesch grade level). The text was divided into three sections, each including a subtopic header (Capturing Greenhouse Gases, A New Approach in Norway, and Underground or Underwater). Collectively, the three sections discussed the problem and possible solutions. The sections were related to one another and contained information that was intended to be integrated.

In both experiments, all participants were forewarned that they would be asked to type in everything they could remember from the studied material. Participants in both groups were then given 4–5 min (depending on text section length) to study each section, one at a time. In the interim-test group, immediately after studying Section 1, participants were shown the subtopic header with an empty text field for them to use to type in everything they could remember from

the section. Participants were given 5 min, after which the computer automatically advanced participants to study Section 2, and so on until each section had been studied and recalled.² In the no-interim-test group, participants studied all three sections before any testing took place. After Section 3, participants were shown the subtopic header for Section 3 with an empty text field and were given 5 min to type in everything they could remember from that section. Recall of Sections 1 and 2 was then collected in a similar manner. Importantly, recall of the target Section 3 took place directly after study of Section 3 for both the interim-test and no-interim-test groups.

Results and discussion

For scoring, text sections were parsed into idea units corresponding to the content of a simple phrase. Credit was assigned for verbatim responses or correct paraphrases. Recall of the target Section 3 was of greatest interest (recall of Sections 1 and 2 is reported in the [Appendix](#), for archival purposes). Mean recall for Section 3 in each group is reported in [Fig. 1](#). Section 3 recall was significantly greater for the interim-test group than for the no-interim-test group in [Experiments 1A and 1B](#) [$t(127) = 5.55, p < .001$, and $t(38) = 2.20, p = .034$]. Thus, our results extend Szpunar et al.'s (2008) findings with word pairs by showing that interim testing also facilitates learning of subsequent text material.

To what extent was the enhanced recall of Section 3 in the interim-test group due to release from proactive interference from the content of Sections 1–2? Paralleling the analyses reported by Szpunar et al. (2008), for each participant we computed the number of idea units from Sections 1 and 2 included in recall of Section 3. We found the same qualitative pattern of intrusions as had Szpunar et al. In fact, not a single participant in the interim-test group in either experiment had an intrusion from a preceding text section. In contrast, intrusions in the no-interim-test group were significantly greater than zero in [Experiment 1A](#) ($M = 1.5, SE = 0.3, t(64) = 4.46, p < .001$, and in [Experiment 1B](#) ($M = 1.3, SE = 0.6, t(18) = 2.26, p = .037$). However, overall intrusion rates were relatively low compared to those reported by Szpunar et al., suggesting that release from proactive interference may play less of a role in interim-test effects with complex material than with word lists.

² A methodological difference between the present study and Szpunar et al. (2008) is the exclusion of intervening activity (i.e., math problems) before immediate recall. Szpunar et al. included 1 min of math problems as a means to clear working memory prior to recall. However, this is less of a concern with lengthier text sections, because only a relatively small amount of an entire section remains in working memory, and learners tend to output text in canonical order rather than demonstrating a recency effect.

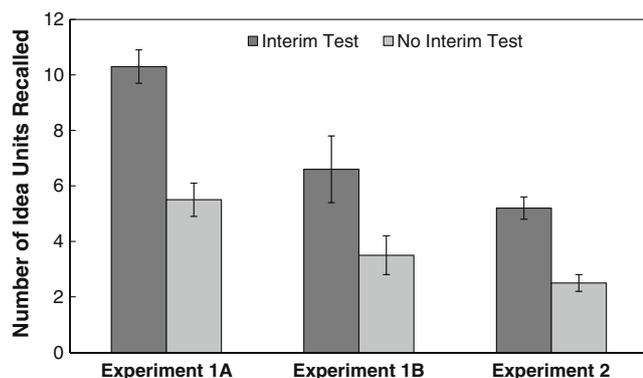


Fig. 1 Mean number of idea units correctly recalled from the target section for each group in Experiments 1A, 1B, and 2 (different text materials were used in each experiment). Error bars represent standard errors

Experiment 2

Experiments 1A and 1B suggest that interim tests over prior text material facilitate learning of subsequent material. An alternative explanation is that the two groups did not differ in learning but rather in output. Individuals in the no-interim-test group may have learned Section 3 just as well as those in the interim-test group, and perhaps even better, to the extent that reading text sections in an uninterrupted fashion may have facilitated the integration of information across all sections. If so, learners in the no-interim-test group may have had greater difficulty discriminating which information came from Section 3 (vs. Section 1 or 2). Due to uncertainty of sources, learners in this group may have engaged in more stringent output monitoring, and thus may not have reported all of the information they learned for Section 3. Furthermore, recall of sections in the interim-test group took place in the canonical 1–2–3 order, whereas recall in the no-interim-test group was prompted in a noncanonical 3–1–2 order. This may have further impaired recall in the no-interim-test group if learners formed a coherent, integrated representation of the text. To minimize these potential disadvantages to the no-interim-test group, in Experiment 2, this group completed unconstrained free recall for information from all text sections after studying Section 3. Thus, learners did not have to discriminate which information came from each text section and could recall the material in canonical order.

Method

Participants and design Undergraduates from Kent State University who participated for course credit were randomly assigned to one of two groups (interim test and no interim test; $n_s = 54$ and 59 , respectively).

Materials and procedure The materials included an expository text on inconsistencies between Hollywood's depiction of history and factual history (1,319 words, 11.4 Flesch grade level), divided into four sections. The sections discussed why filmmakers choose to modify historical facts and then described a specific example of a film in which this was done, comparing how a historical event was portrayed in the film with what had occurred in actuality. Hence, the text sections contained information that was intended to be integrated.

The procedure was the same as in Experiments 1A and 1B, except that participants in the no-interim-test group read all of the text sections and then were given 20 min for free recall of information from all four sections.

Results and discussion

Mean recall for Section 4 in each group is reported in Fig. 1. Section 4 recall was significantly greater for the interim-test group than for the no-interim-test group, $t(111) = 5.62$, $p < .001$. Even with the change in procedure to reduce the task demands for output monitoring and to afford canonical recall order, recall was still significantly lower for the no-interim-test group.

Experiment 3

Experiment 2 further established that interim tests over prior text material facilitate learning of subsequent new material. However, an alternative interpretation is that the effects were not due to interim testing per se, but rather to intervening activity. Experiment 3 addressed this possibility by including a group that completed math equations between text sections. If intervening activity facilitates learning, recall for the interim-math group will resemble that for the interim-test group. Conversely, if interim testing facilitates learning, the interim-test group will outperform the interim-math group.

Method

Participants and design Undergraduates from Kent State University who participated for course credit were randomly assigned to one of three groups (interim test, no interim test, and interim math; $n_s = 29$, 29 , and 30 , respectively).

Materials and procedure The materials included the text used in Experiment 1B. The procedure for the interim-test and no-interim-test groups was the same as in Experiments 1A and 1B, except that the time allotted for recall was changed to 5–6 min (depending on section length) due to a

concern that participants in the interim-test group were being stopped before they had completed recall for some sections. Participants in the interim-math group solved math problems for 5 min between study of Sections 1 and 2 and between Sections 2 and 3. Recall of Section 3 took place immediately after studying Section 3, as in Experiments 1A and 1B (followed by recall of Sections 1 and 2). Thus, for all three groups, recall for the target Section 3 took place immediately after studying Section 3.

Results and discussion

Mean recall for Section 3 in each group is reported in Fig. 2. A one-way ANOVA revealed a significant main effect, $F(2, 85) = 7.51$, $MSE = 14.37$, $p = .001$. Replicating our previous findings, Section 3 recall was greater for the interim-test group than for the no-interim-test group, $t(56) = 3.41$, $p = .001$. Furthermore, Section 3 recall was greater for the interim-test group than for the interim-math group, $t(57) = 2.91$, $p = .005$. These results indicate that interim testing in particular facilitates learning, rather than intervening activity in general.

Concerning the extent to which this effect reflects release from proactive interference, as in Experiments 1A and 1B, not a single participant in the interim-test group had an intrusion from Section 1 or 2 during Section 3 recall. However, intrusions were also infrequent in the no-interim-test group ($M = 0.2$, $SE = 0.1$), $t(28) = 1.80$, $p = .083$, and the interim-math group ($M = 0.4$, $SE = 0.2$), $t(29) = 1.76$, $p = .09$.

Experiment 4

Experiment 3 further established the interim-test effect for text and provided evidence that the effect is not due to

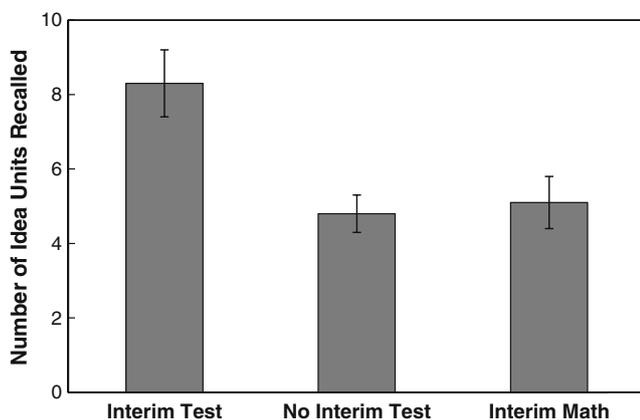


Fig. 2 Mean number of idea units correctly recalled from the target section for each group in Experiment 3. Error bars represent standard errors

intervening activity per se. One goal of Experiment 4 was to evaluate another explanation for interim-test effects concerning test expectancy. Although all participants received instructions about the free recall test prior to studying, recall of Sections 1 and 2 prior to Section 3 in the interim-test group may have resulted in a better expectation of what type of recall test to expect. To evaluate this possibility, Experiment 4 included a *practice-test* group, in which participants read and recalled a short, unrelated text before studying the target material, to illustrate the type of test they should expect to take after reading all sections of the target text. To the extent that interim testing produces an effect above and beyond any effect of test expectancy, performance would be greater in the interim-test group than in the practice-test group.

The second goal of Experiment 4 was to more directly evaluate the potential contribution of release from proactive interference to the interim-test effect. Experiment 4 included a *Section-3-only* group, in which participants were not exposed to Sections 1 and 2 prior to reading and recalling Section 3. If release from proactive interference plays a minimal role in the interim-test effect with text (as suggested by the overall low intrusion rates in Experiments 1A, 1B, and 3), performance in the Section-3-only group would be lower than in the interim-test group, and may not even differ from performance in the no-interim-test group.

Method

Participants and design Undergraduates from Kent State University who participated for course credit were randomly assigned to one of four groups (interim test, no interim test, practice test, and Section 3 only; $n_s = 23, 26, 27$, and 26, respectively).

Materials and procedure The materials included the first three sections of the text used in Experiment 2. The procedure for the interim-test and no-interim-test groups were the same as in Experiment 3. In the practice-test group, participants first read and then immediately recalled a short passage about silkworms, to ensure that participants had clear expectations for what type of recall test would be administered for the target material. After the practice test, the procedure was the same as in the no-interim-test group. In the Section-3-only group, participants were given the allotted time to read and then immediately recall Section 3, without prior study of Sections 1 or 2.

Results and discussion

Mean recall for Section 3 in each group is reported in Fig. 3. A one-way ANOVA revealed a significant main

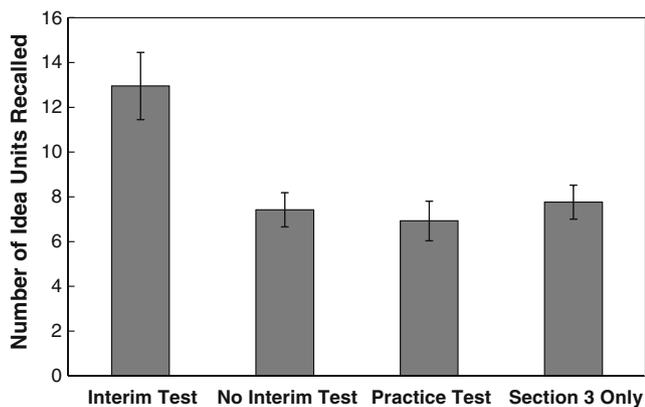


Fig. 3 Mean number of idea units correctly recalled from the target section for each group in Experiment 4. Error bars represent standard errors

effect, $F(1, 98) = 7.57$, $MSE = 24.94$, $p < .001$. Once again, Section 3 recall was greater for the interim-test group than for the no-interim-test group, $t(47) = 3.40$, $p = .001$. Importantly, Section 3 recall was also greater for the interim-test group than for the practice-test group, $t(48) = 3.58$, $p < .001$, indicating that increased recall was not the result of test expectancy. One might have expected the practice test to also have produced some facilitative effect on learning of the target section. However, there was no memorial benefit, given that two other, untested text sections intervened between the practice test and study of Section 3. These findings are similar those reported by Szpunar et al. (2008, Exp. 2), showing that as the number of previously untested lists increased, correct recall of List 5 decreased.

Concerning the role of proactive interference, Section 3 recall was greater for the interim-test group than for the Section-3-only group, $t(47) = 3.19$, $p = .002$, indicating that increased recall was not the result of reduced proactive interference. Once again, not a single participant in the interim-test group had an intrusion from an earlier section. Intrusions were marginally greater than zero in the no-interim-test group ($M = 0.8$, $SE = 0.4$), $t(25) = 2.01$, $p = .056$, and significantly greater than zero in the practice group ($M = 0.8$, $SE = 0.2$), $t(26) = 3.81$, $p = .001$, although intrusion rates were relatively low overall.

General discussion

Five experiments demonstrated that interim testing of prior text material facilitates learning of subsequent new material. These results extend beyond those of Szpunar et al.'s (2008) study by showing that interim-test effects generalize to complex text material. The results of

Experiments 1A, 1B, 3, and 4 are particularly striking—although all groups recalled Section 3 immediately after reading it, the interim-test groups consistently recalled nearly twice as much information as the other groups. Experiment 2 addressed the possible consequences of output monitoring and noncanonical recall for the no-interim-test group, but the interim-test group continued to show a marked advantage. Experiments 3 and 4 further established that the observed recall advantage was due to interim testing in particular, rather than to intervening activity more generally or to test expectancy differences.

How does interim testing enhance learning of subsequent text material? Szpunar et al. (2008) suggested that one mechanism underlying the effect for word lists involves release from proactive interference. Given their significant reductions in intrusions of prior list items during recall of List 5 following interim tests versus no interim tests, Szpunar et al. concluded that interim tests made it easier for participants to discriminate which words came from each list.

To what extent does release from proactive interference underlie the interim-test effects observed in the present experiments? Although we observed the same qualitative pattern of intrusions as Szpunar et al. (2008), intrusions in the no-interim-test groups were relatively infrequent (typically less than one idea unit from any of the preceding sections). Furthermore, Experiment 4 showed that eliminating proactive interference from preceding sections did not enhance target section recall. Therefore, proactive interference appears to play less of a role in the interim-test effects with text materials than with word lists.

Another possible explanation for facilitated learning involves the opportunity for additional study of any recalled information. However, substantial research on testing effects has shown that testing yields advantages beyond reexposure to study material (Roediger & Butler, 2011). Furthermore, Szpunar et al. (2008, Exp. 3) included a group with restudy trials after Lists 1–4, and recall for List 5 was significantly higher for participants who took interim tests versus restudying. These results suggest that the interim-test effect is due to testing rather than to reexposure to content from preceding sections prior to recall of the target section, although a direction for future research would be to further explore the effects of testing relative to other kinds of intervening activity.

What other mechanisms might underlie interim-test effects with text? One possible explanation is that an interim test over a preceding section improves memory for that information, such that the information is more readily accessible when reading a subsequent section. Enhanced accessibility of prior information might facilitate compre-

hension of subsequent material when reading connected discourse (as in Exps. 1B–4). How enhanced accessibility would facilitate learning of topically related but otherwise separate texts (Exp. 1A) is less clear, unless information could be used in a compare–contrast manner.

Another possibility is that retrieval may engender the use of more effective encoding strategies and/or mediators. This idea is related to a distinction in the testing effect literature between direct and mediated effects of testing (Roediger & Karpicke, 2006). *Direct effects* refer to memorial benefits of testing that arise from the act of taking a test itself (e.g., retrieving information from memory increases the memorability of that information). *Mediated effects* refer to memorial benefits that are due to an influence of testing on subsequent encoding or study behavior. For example, Pyc and Rawson (2010) reported that retrieval practice with word pairs can lead to the development of more effective mediators during subsequent restudy. Bahrick and Hall (2005) suggested that the experience of retrieval failure may be particularly important for encouraging learners to shift to more effective encoding strategies. Similar processes may be at work in the present research, such that the difficulty of retrieving information during an interim test over Material A could enhance encoding strategies during study of Material B.

To conclude, the present results establish that interim tests over initial text material can enhance the learning of subsequent text material. Concerning the educational implications of this finding, students may profit from the use of interim self-testing while reading lengthy textbook chapters or in-between studying sets of notes for different classes. Additionally, teachers could administer an informal recall test over the content of one topic before moving on to the next topic within a lecture. However, these are only tentative suggestions at this point, as further research will be needed to explore the generality of interim-test effects. One possible direction would be to evaluate the extent to which these effects depend on variables such as the motivational level of students. Further exploration of the specific mechanisms that underlie these effects will also be an important direction for future research.

Author note Mary A. Pyc is now at the Department of Psychology, Washington University in St. Louis.

The research reported here was supported by supported by a Collaborative Award from the James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind and Behavior, and by the Institute of Education Sciences, U.S. Department of Education, through Grants R305H050038 and R305A080316 to Kent State University. The opinions expressed are those of the authors and do not represent views of the Institute or of the U.S. Department of Education. Thanks to Nicole Gonzalez, Phil Grimaldi, Caitlin Metelko, and Sara Smith for assistance with data collection.

Appendix

Table 2 Mean number of idea units recalled (with standard errors) for nontarget sections

	Section 1	Section 2	Section 3
Experiment 1A			
Interim test	9.7 (0.6)	9.8 (0.7)	N/A
No interim test	4.9 (0.5)	3.4 (0.5)	N/A
Experiment 1B			
Interim test	6.9 (0.9)	7.2 (1.1)	N/A
No interim test	3.6 (0.9)	4.2 (1.2)	N/A
Experiment 2			
Interim test	7.3 (0.4)	5.4 (0.4)	6.9 (0.6)
No interim test	5.5 (0.4)	2.8 (0.3)	3.8 (0.4)
Experiment 3			
Interim test	9.0 (0.9)	10.1 (1.0)	N/A
No interim test	4.4 (0.6)	3.5 (0.6)	N/A
Interim math	4.1 (0.6)	3.9 (0.7)	N/A
Experiment 4			
Interim test	14.2 (1.6)	12.5 (1.1)	N/A
No interim test	8.8 (0.8)	7.4 (0.9)	N/A
Practice test	8.8 (1.3)	6.5 (1.2)	N/A

References

- Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*, 566–577. doi:10.1016/j.jml.2005.01.012
- Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*, 153–170. doi:10.1016/j.jml.2009.04.004
- Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, *18*, 49–57. doi:10.1080/09658210903405737
- Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 431–437. doi:10.1037/0278-7393.33.2.431
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571. doi:10.1037/0096-3445.135.4.553
- Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science*, *20*, 66–73. doi:10.1111/j.1467-9280.2008.02245.x
- Darley, C. F., & Murdock, B. B., Jr. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*, 66–73. doi:10.1037/h0031836

- Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, 8, 200–224. doi:10.1016/0022-2496(71)90012-5
- Karpicke, J. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469–486. doi:10.1037/a0017341
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330, 335. doi:10.1126/science.1191465
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27. doi:10.1016/j.tics.2010.09.003
- Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210. doi:10.1111/j.1745-6916.2006.00012.x
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1392–1399. doi:10.1037/a0013082
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, 13, 181–193. doi:10.1016/S0022-5371(74)80043-5