Journal of Experimental Psychology: Applied
2016, Vol. 22, No. 3, 305–318

1076-898X/16/$12.00 http://dx.doi.org/10.1037/xap0000087

Interpolated Testing Influences Focused Attention and Improves Integration of Information During a Video-Recorded Lecture

Helen G. Jing
Harvard University

Karl K. Szpunar
University of Illinois at Chicago

Daniel L. Schacter
Harvard University

Although learning through a computer interface has become increasingly common, little is known about how to best structure video-recorded lectures to optimize learning. In 2 experiments, we examine changes in focused attention and the ability for students to integrate knowledge learned during a 40-min video-recorded lecture. In Experiment 1, we demonstrate that interpolating a lecture with memory tests (tested group), compared to studying the lecture material for the same amount of time (restudy group), improves overall learning and boosts integration of related information learned both within individual lecture segments and across the entire lecture. Although mind wandering rates between the tested and restudy groups did not differ, mind wandering was more detrimental for final test performance in the restudy group than in the tested group. In Experiment 2, we replicate the findings of Experiment 1, and additionally show that interpolated tests influence the types of thoughts that participants report during the lecture. While the tested group reported more lecture-related thoughts, the restudy group reported more lecture-unrelated thoughts; furthermore, lecture-related thoughts were positively related to final test performance, whereas lecture-unrelated thoughts were negatively related to final test performance. Implications for the use of interpolated testing in video-recorded lectures are discussed.

Keywords: interpolated testing, online learning, video-recorded lectures, mind wandering, information integration

Classroom lectures are a central feature of a college student's educational experience. With the advent of online learning, flipped classrooms, and massive open online courses, all of which use video-recorded lectures (Breslow et al., 2013), the need to develop instructional techniques that can serve to improve the quality of learning from lectures represents an important research endeavor (Schacter & Szpunar, 2015). It is important to note that the temporally extended nature of lectures makes them ideal for studying learning in that patterns of learning and attention to material may shift over long sequences of study. In the present article, we examine changes in learning in the form of information integration and focused attention during a video-recorded lecture.

This article was published Online First June 13, 2016.

Helen G. Jing, Department of Psychology, Harvard University; Karl K. Szpunar, Department of Psychology, University of Illinois at Chicago; Daniel L. Schacter, Department of Psychology, Harvard University.

This research was supported by the Spark Grant awarded by the Harvard Initiative for Learning and Teaching. We thank Nicholas Christakis for graciously providing the lecture video and lecture materials. We also thank Jenny Beizer and Stella Tu for assistance with various aspects of the experiments.

Correspondence concerning this article should be addressed to Helen G. Jing, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: hjing@fas.harvard.edu

Learning and Information Integration

There is a vast literature demonstrating that the process of retrieving information from memory during testing can enhance learning, a phenomenon known as the *testing effect* (for review, see Roediger & Butler, 2011; Roediger & Karpicke, 2006). A subset of this literature has studied the effect of interspersing a longer study episode with retrieval practice in the form of free recall or short quizzes (i.e., interpolated testing), in contrast to testing after completion of a study episode. The use of an interpolated testing paradigm not only allows one to examine direct effects of testing (i.e., how testing improves retention of material) but also allows for studying indirect effects of testing, such as how retrieval during interpolated quizzes can enhance learning of subsequently presented material (for review, see Pastötter & Bäuml, 2014). In general, the benefits of interpolated testing on learning and retention exceed those of additional study and have been demonstrated in a variety of educationally relevant materials, including but not limited to word lists (Szpunar, McDermott, & Roediger, 2008), prose passages (Wissman, Rawson, & Pyc, 2011), classroom settings (e.g., McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), and video-recorded lectures (e.g., Szpunar, Khan, & Schacter, 2013).

One important feature of learning concerns the need for students to process relatively large amounts of information that are presented over extended periods of time. Doing so involves both *item-specific processing* (i.e., encoding single facts or units of information) and *relational processing* (i.e., encoding and integrat-

ing the relations between separate pieces of information; e.g., Hunt & Einstein, 1981; Hunt & McDaniel, 1993); indeed, recent research has suggested that combining both item-specific and relational processing enhances learning (Hunt, 2013; Hunt & Rawson, 2011). However, although testing is known to enhance item-specific memory, there exists conflicting evidence for how testing may affect relational processing or integration, on which we elaborate in the following paragraphs.

Within a laboratory paradigm that presents information across separate learning sequences (e.g., separate word lists, text passages, or lecture segments), integration of learned material may involve two key features: integration of material within each consecutive sequence, and integration of related material across different sequences. First, experimental studies using word list stimuli have shown that introducing interpolated tests between the presentation of consecutive word lists enhances the discriminability of each list during recall. For example, Szpunar et al. (2008) presented participants with five separate word lists, and some participants were required to recall each list prior to learning the next list (tested group) whereas other participants either completed a distracting task following the first four lists in the study sequence or restudied those first four lists but always recalled words from the fifth list (nontested groups). During list five recall, tested participants produced fewer intrusions of words from previous lists than nontested participants, suggesting that those who were tested were better able to segregate the recall of words that originated from different lists. Moreover, Chan and McDermott (2007) showed that interpolated testing following study of different word lists enhanced participants' memory for which list a word originated from, suggesting that the effects of testing on item discrimination persist beyond the original study episode (see also Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011). These measures of discriminability are typically interpreted as evidence for segregation of learned material. However, between-list segregation may also serve as an important measure of within-list integration, particularly when information within a list is highly related. For instance, in the context of lectures, wherein information presented within any given lecture segment is likely to be highly related, it would be important to know whether interpolated tests serve to support the integration of information presented within segments.

Second, prior work has also examined the effect of interpolated testing on integration of related knowledge learned across extended sequences of study, although evidence has thus far been mixed (Peterson & Mulligan, 2013; Rawson, Wissman, & Vaughn, 2015; Wissman & Rawson, 2015; Zaromb & Roediger, 2010). Some researchers suggest that testing should boost integration because testing enhances the accessibility of prior material, which may facilitate the creation of associations between related information presented across a learning sequence. For instance, Zaromb and Roediger (2010; see Experiment 2) asked four groups of participants to study three lists of words that shared overlapping semantic categories. Of particular interest were two groups of participants, one group that studied each list twice before moving onto the next list and one group that studied each list once and recalled that list before moving onto the next list. Following a 1-day retention interval, participants were asked to write down as many words as they could from all three lists in any order. The authors found that participants in the tested group were more likely to cluster their recall across semantic categories that were common across word lists as compared to participants in the nontested group, suggesting that testing supported integration of related materials learned at different points in time (see also, Masson & McDaniel, 1981). More recently, Wahlheim (2015) showed that interpolated testing can help to overcome the occurrence of proactive interference (i.e., when previously learned material interferes with the learning of new material) in the context of an A−B, A−D paired associate learning paradigm via an integration-based mechanism. Specifically, the act of testing A−B pairs (e.g., knee-bone) can facilitate as opposed to interfere with learning of A−D pairs (e.g., knee-bend) because the interpolated tests increase the probability that participants can consciously recollect the change in the cue−response pairing across the experiment. In other words, an interpolated test fosters the formation of an integrated memory trace that helps the participant to keep track of the temporal relationship of the cue-response pairings (Wahlheim, 2015; see also Wahlheim & Jacoby, 2013). Taken together, it appears that testing can facilitate knowledge integration across extended study sequences.

Alternatively, researchers have hypothesized that testing may impair integration or relational processing (relative to restudying material) because testing tends to direct one's attention toward item-specific processing that may detract from noticing the relations between items (e.g., Peterson & Mulligan, 2013) or because testing a specific portion of material may disrupt the individual from connecting information from different sections (e.g., Wissman & Rawson, 2015). For example, Peterson and Mulligan (2013) presented participants with pairs of words (e.g., moon—spoon) in which target words were drawn from six broader taxonomic categories (e.g., animals, fruits, metals). After the initial presentation phase, participants were either asked to restudy the word pairs or retrieve the target words in response to a presented cue (e.g., retrieve "spoon" when presented with "moon"). In contrast with the studies presented earlier, these researchers found that during a free recall test, participants in the tested group showed less clustering by taxonomic category relative to the restudy group, and interpreted these data as evidence that testing impairs relational processing between items within these broader categories. However, Rawson et al. (2015) failed to replicate this effect. In another set of studies, Wissman and Rawson (2015) presented participants with text passages, in which the content of each passage was highly interrelated. When cued with a sentence from the text and asked to type in other information from the text related to the cue, data from two experiments (Wissman & Rawson, 2015; see Experiments 6 and 7) indicated that participants who read through the whole text in an uninterrupted fashion did not integrate any more information than those who regularly recalled information from the separate text segments. In fact, the aggregate data suggested that those who regularly recalled information between text segments connected more information from other parts of the text than the whole-text group (particularly for content derived from the same segment as the cue). Thus, these data do not support the authors' hypothesis that testing should impair the integration of related ideas studied in the context of more complex materials. Given the conflicting evidence presented above, whether interpolated testing may serve to enhance the relational integration of knowledge both within and across different segments of other

materials, such as a video-recorded lecture, remains an open and important question.

Mind Wandering

Another salient aspect of learning concerns the ability to sustain focused attention for extended periods of time, and scholars have long noted that students experience difficulty in doing so during lectures (Bunce, Flens, & Neiles, 2010; Johnstone & Percival, 1976; Wilson & Korn, 2007). The study of shifts in focused attention in the classroom (i.e., shifts away from task-related thoughts; TRT) has been operationalized in terms of task-unrelated thoughts (TUT) or mind wandering, the tendency for the contents of one's mind to drift away from an ongoing task toward unrelated inner thoughts and feelings (e.g., daydreaming, thinking about one's personal past or future), as well as task-related interferences (TRI), the experience of interfering thoughts related to appraisals of an ongoing task (e.g., task difficulty or length; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Smallwood et al., 2004; Smallwood & Schooler, 2006). Much existing research has focused on *unintentional mind wandering* (i.e., when one's mind spontaneously drifts to task-unrelated thoughts despite fully intending to pay attention), which is thought to reflect either a demand in executive control resources during mind wandering that detracts from the primary task (Smallwood, 2010), or failures in executive control (McVay & Kane, 2010). However, researchers have also highlighted instances of *intentional mind wandering*, when one deliberately thinks about other things unrelated to an ongoing task (Giambra, 1995; Seli, Carriere, & Smilek, 2015; Seli, Cheyne, Xu, Purdon, & Smilek, 2015). Overall, studies have shown that mind wandering is a common occurrence during lectures both within the classroom (Bunce et al., 2010; Johnstone & Percival, 1976; Lindquist & McLean, 2011) and in online or video-recorded settings (Hollis & Was, 2016; Khan, 2012; Koller, 2011; Risko, Anderson, Sarwal, Engelhardt, & Kingstone, 2012). Moreover, the occurrence of mind wandering in these contexts is negatively associated with learning outcomes (Hollis & Was, 2016; Lindquist & McLean, 2011; Risko et al., 2012).

In an effort to overcome the negative impact of mind wandering on learning in lecture-based contexts, Szpunar, Khan, and Schacter (2013) showed that interpolating brief tests within a video-recorded lecture could reduce the frequency of mind wandering, such that students who were tested periodically during the lecture were less likely to mind wander than students who did not receive periodic tests. It is important that the reduction in mind wandering was related to increases in learning (see also, Szpunar, Jing, & Schacter, 2014). Although these preliminary findings are encouraging, our prior work focused on a single, 20-min introductory statistics lecture, leaving open the question of whether similar effects can be observed with other lectures. Furthermore, in these experiments, mind wandering was broadly defined as instances where participants' attention strayed away from the lecture content. During a lecture that covers material perceived as difficult to master or perhaps even boring, such as an introductory statistics lecture (Gal & Ginsburg, 1994), one might expect that the majority of participants' supposedly "off-task" thoughts are likely task-unrelated (e.g., daydreaming; see Smallwood et al., 2003). However, using other lecture materials, such as a more interesting lecture or one in which the material is highly interrelated across different segments, may result in a different profile of thoughts, including those that are more relevant to the task at hand. Hence, more work is needed to assess the extent to which the benefits of interpolated testing on mind wandering and learning generalize to other materials, including materials that students may find inherently more engaging or interesting than statistics.

The Present Studies

To address the extent to which the benefits of interpolated testing on learning extend to other contexts and measures of learning, we conducted two experiments in which students were asked to learn from a 40-min video-recorded lecture on public health in the United States. In Experiment 1, we first assessed the extent to which the benefits of our interpolated testing manipulation, such as reduced mind wandering and improved learning, generalize to other types of lecture content. Second, the highly interrelated content of the lecture is conducive to assessing the extent to which students were able to integrate knowledge from across the entire lecture. In Experiment 2, we further probed different types of thoughts participants reported while watching the video-recorded lecture and examined how these thoughts related to learning.

Experiment 1

Method

Participants. A total of 39 healthy undergraduate students (ages 18 to 25, $M = 20.4$ years, 21 female) were recruited from Harvard College and Boston University. Participants were paid or received course credit for their participation. All participants had normal vision and no history of neurological or psychological impairment. Prior to data collection, we conducted a power analysis using the effect size ($d = 1.05$) reported by Szpunar et al. (2013) for the difference in mind wandering rates in the tested and restudy groups. This analysis indicated that using a sample size of 16 participants per group would allow us to detect effects of interpolated testing on mind wandering frequency with power equal to or greater than 0.80. Data collection was stopped once it was determined that approximately enough usable participants had been run to reach this number. A total of 3 participants were excluded due to technical difficulties during the experiment, leaving 36 participants in the final sample.

Lecture and study materials. We used a 40-min sociology lecture video on the role of medicine and other social factors on public health in the 20th century (Sociology 190, Department of Sociology, Harvard University). The video was divided into eight 5-min segments using iMovie software. Participants watched the lecture video and completed other experimental tasks using Qualtrics survey software.

Design and procedure. The experimental session lasted a total of 2 hr and involved (a) watching a 40-min lecture interspersed with mind wandering probes and other activities and (b) a two-part final cumulative test. See Figure 1 for a diagram of the experimental procedure.

Lecture. Participants were informed that the video-recorded lecture would be divided into eight segments of equal length and that they would complete a number of tasks between each segment.

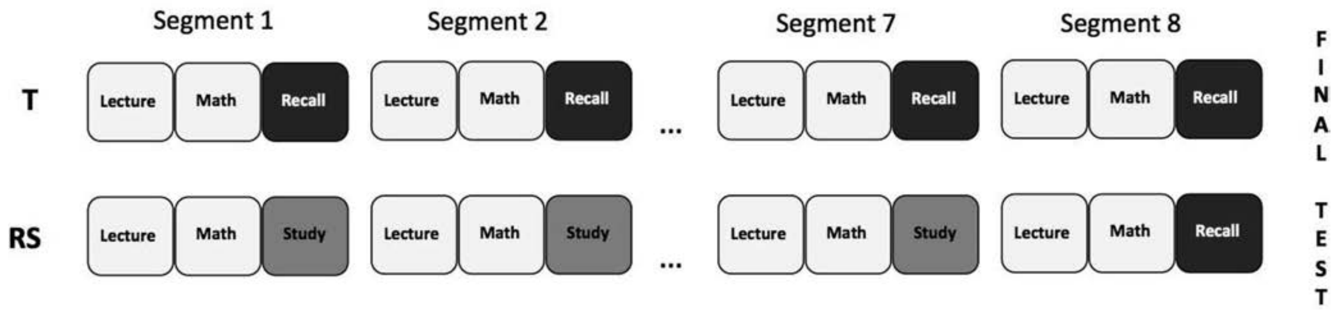

Figure 1. Schema of experimental design for tested (T) and restudy (RS) groups.

As they watched the lecture, they were provided with a packet of lecture slides that corresponded to the material presented on the computer screen, and were told they could use the slides in any way that might help them learn from the lecture. After each 5-min lecture segment, participants completed 1 min of math problems that were unrelated to the content of the lecture (e.g., $12 \times 7 = ?$); six problems were presented and participants were given 10 seconds to answer each problem. Participants engaged in solving math problems after each segment to minimize retrieval of information from working memory during free recall.

Following the math problems, participants had the opportunity to engage in one of two activities for 2.5 min: They either (a) freely recalled information that they learned from the previously watched lecture segment on a lined sheet of paper or (b) studied the lecture slides and any notes they took from the previous lecture segment. It is important to note that participants were informed that a computer program would randomly determine whether they could recall or restudy information after any given segment, such that they could be asked to engage in either activity after every lecture segment, after none of the segments, or anywhere in between. Participants were told that they would later receive a final cumulative test that would assess their knowledge about the whole lecture, regardless of the activities they completed during the lecture. In reality, half of participants were given the opportunity to freely recall information after all eight segments of the lecture (tested group), and half of participants studied the lecture slides for the first seven segments and recalled information following only the eighth segment of the lecture (restudy group). Upon finishing the lecture, participants rated their interest in the lecture and perceived difficulty in comprehending the lecture material (both on a scale of 1 to 10) and were subsequently given a 5-min break during which they worked on a number search puzzle.

Mind wandering probes. Participants were told that a visual cue indicated by the phrase "Are you mind wandering?" would appear on the computer screen at random points during the lecture, and that they should respond *yes* or *no* on the computer screen whenever they saw this cue. They were instructed to respond *yes* if their attention strayed from the lecture content (i.e., if they were thinking about something other than what the lecturer was discussing during the probe's appearance). The visual cue remained on the screen for 5 s as the lecture continued, and was accompanied by a short auditory cue (i.e., a bell) to ensure that participants noticed the cue. Four different mind wandering probe sequences were used in the study. For each sequence, the mind wandering probe occurred at a randomly determined time point during each segment that was at least 30 s into the segment and 30 s before the segment was complete. For instance, one sequence involved probes that occurred 155 s, 221 s, 49 s, 95 s, 112 s, 203 s, 138 s, and 76 s into the first through eighth segments, respectively. The presentation of these mind wandering probe sequences was counterbalanced across participants.

Final cumulative test. Participants were administered the final cumulative test in two parts. In the first part of the final test, participants were given 7 min to freely recall as much information as they could remember from the entire lecture (we refer to this portion as *free recall*). Following free recall, participants were presented with images of 16 out of the total 34 lecture slides that were viewed during the lecture; two slides were selected from each segment of the lecture, and each slide was presented individually on the computer screen (we refer to this portion as *cued recall*). Participants were given 1 min to elaborate upon the content of each lecture slide, and were also encouraged to elaborate upon how the individually presented slide related to slides from other parts of the lecture. For both parts of the final test, participants typed the information they recalled in a text box provided on the computer screen.

Note taking and free recall scoring. To assess note taking, we examined the proportion of lecture slides on which students took additional notes. We used this measure as a rough indication of student engagement to examine whether interpolated testing influenced note-taking behavior across the tested and restudy groups.

The 40-min lecture video was transcribed and participants' free recall responses were compared to the original lecture content to check for accuracy. We scored participants' free recall and cued recall responses for the total amount of factually correct information that was recalled to assess overall performance on the final cumulative test. Each unit of factually correct information was scored to be worth one point. For example, "The measles vaccine was introduced after measles mortality rates had largely declined" was worth one point, and "Measles and scurvy mortality rates declined in parallel, although treatments for measles had no effect on scurvy and vice versa" was worth two points (i.e., one point for the "measles and scurvy mortality rates declined in parallel," and one point for "treatments for measles had no effect on scurvy and vice versa"). Participant performance on each portion of the final test was computed as the summed total of all recalled information that was verified to be factually correct.

Although all participant responses were scored according to the transcribed lecture content for accuracy, two raters were trained to score free recall and cued recall responses to assess interrater reliability. Both raters were blind to condition and separately scored 5 participants' responses from both tests. High interrater reliability was obtained for the units of information recalled (standardized Cronbach's $\alpha = .986$). Rater 1 scored 56% of participant responses, and Rater 2 scored 44% of participant responses.

Integration. We assessed the degree to which participants integrated learned material from the lecture during the final test in two ways. First, we examined integration of related information that was learned within each individual lecture segment during free recall. We used the adjusted-ratio-of-clustering method (ARC; Roenker, Thompson, & Brown, 1971) to assess group differences in the clustering of free recall output according to lecture segment; that is, the frequency with which participants followed a piece of information recalled from one lecture segment with another item recalled from the same segment, as opposed to information from a different lecture segment. By this method, higher ratios of clustering would indicate greater integration of information that was learned within each individual lecture segment.

Second, we also examined group differences in the degree to which participants integrated related material across different segments of the lecture. As previously mentioned, during cued recall, all participants were encouraged to elaborate upon information about each individually presented lecture slide on the computer screen for 1 min. Critically, participants were also encouraged to elaborate upon how the presented slide related to slides from other parts of the lecture. We operationalized this novel measure of integration as the amount of information that participants recalled that was not directly derived from the presented slide, but rather was mentioned in relation to other lecture slides. Once again, each unit of factually correct, integrated information was scored to be worth one point, and participants' integration scores were computed as the summed total amount of information that was recalled from other slides in relation to the presented slide.

Results

Demographics. The tested and restudy groups did not differ in their levels of education, $t(34) = 1.27$, $p > .2$, 95% CI = [−.33, 1.44], $d = 0.42$, interest in the lecture, $t(34) = .35$, $p > .5$, 95% CI = [−.87, 1.23], $d = 0.12$, or perceived difficulty in comprehending lecture material, $t(34) = .55$, $p > .5$, 95% CI = [−.58, 1.01], $d = 0.18$.

Mind wandering and note taking. We conducted a series of two-tailed *t* tests to assess whether the tested and restudy groups differed in the proportion of mind wandering (i.e., proportion of probes to which participants responded *yes*) and note taking behaviors (i.e., proportion of lecture slides on which participants took additional notes) while viewing the lecture. We found no significant difference in mind wandering between the tested and restudy groups ($M_{\text{tested}} = 20.8\%$, $SE = .04$; $M_{\text{restudy}} = 23.6\%$, $SE = .04$), $t(34) = -.45$, $p > .5$, 95% CI = [−.15, .10], $d = 0.15$ (see Figure 2A). However, the tested group took significantly more notes than the restudy group ($M_{\text{tested}} = 82.2\%$, $SE = .04$; $M_{\text{restudy}} = 66.2\%$, $SE = .06$), $t(34) = 2.11$, $p < .05$, 95% CI = [.01, .31], $d = 0.70$ (see Figure 2B).

Eighth segment recall. Next, to assess whether testing during the first seven segments was associated with improved learning of the final lecture segment, we compared initial recall of information after the eighth segment of the lecture between the tested and restudy groups. The tested group recalled significantly more units of information than the restudy group after the eighth lecture segment ($M_{\text{tested}} = 8.17$ units, $SE = .66$; $M_{\text{restudy}} = 4.56$ units, $SE = .42$), $t(34) = 4.62$, $p < .001$, 95% CI = [2.02, 5.20], $d = 1.54$ (see Figure 3A).

Final test performance. We also assessed performance (i.e., total units of information recalled) on the two portions of the final cumulative test in the tested and restudy groups. First, the tested group recalled significantly more units of information than the restudy group during free recall ($M_{\text{tested}} = 23.94$ units, $SE = 1.83$; $M_{\text{restudy}} = 14.22$ units, $SE = 1.27$), $t(34) = 4.37$, $p < .001$, 95% CI = [5.20, 14.25], $d = 1.46$ (see Figure 3B). Second, the tested group also recalled significantly more units of information than the restudy group during cued recall ($M_{\text{tested}} = 52.17$ units, $SE = 3.07$; $M_{\text{restudy}} = 33.78$ units, $SE = 3.00$), $t(34) = 4.29$, $p < .001$, 95% CI = [9.67, 27.11], $d = 1.43$ (see Figure 3C).

To assess whether improved learning during the eighth lecture segment (as previously assessed by eighth segment recall) persisted over a delay, we compared recall of information during free recall that originated from the eighth segment of the lecture between the tested and restudy groups. The tested group recalled

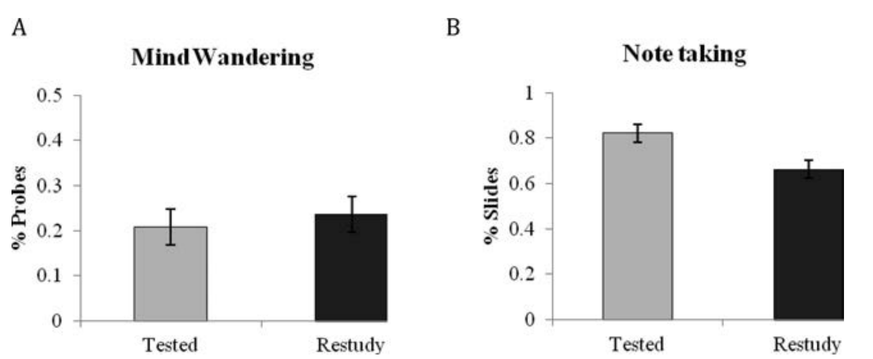

Figure 2. Experiment 1 mean proportions (out of 100%) of (Panel A) mind wandering in response to mind wandering probes and (Panel B) note taking in tested and restudy groups. Error bars represent one standard error of the mean.

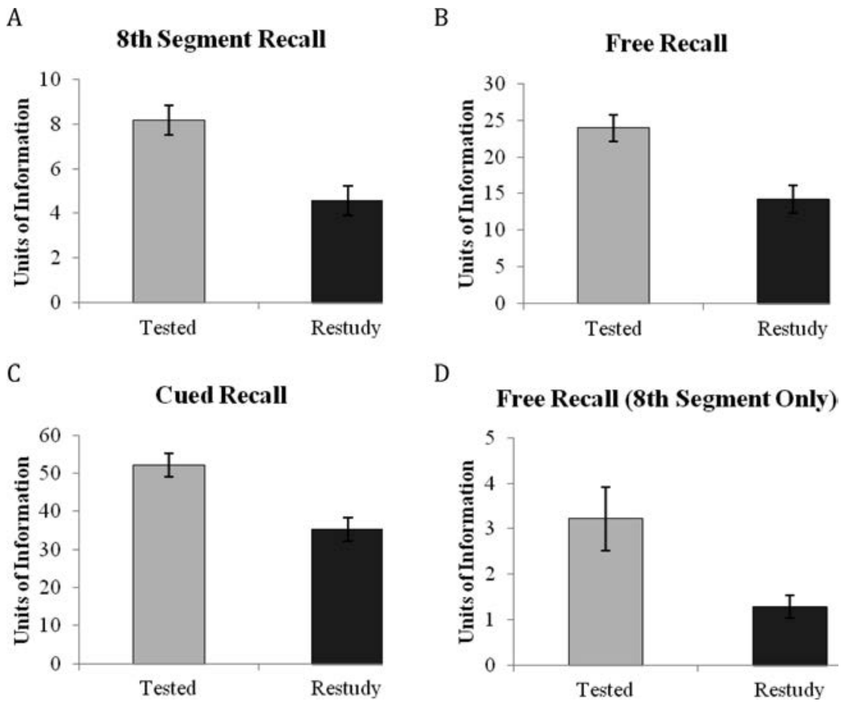

Figure 3. Experiment 1 mean units of information recalled: (Panel A) after the eighth segment of lecture, (Panel B) during free recall, (Panel C) during cued recall, and (Panel D) during free recall that originated from the eighth segment of the lecture. Error bars represent one standard error of the mean.

significantly more units of information from the eighth lecture segment than the restudy group during free recall ($M_{\text{tested}} = 3.22$ units, $SE = .71$; $M_{\text{restudy}} = 1.28$ units, $SE = .25$), $t(21.29) = 2.59$, $p < .001$, 95% CI = [0.38, 3.51], $d = 0.86$ (see Figure 3D). Thus, the benefits of testing on free recall persisted across a short delay.

Integration. During free recall, we found that participants in the tested group ($M_{\text{ARC}} = .81$, $SE = .02$) were more likely than participants in the restudy group ($M_{\text{ARC}} = .68$, $SE = .02$) to cluster their output, $t(34) = 3.77$, $p = .001$, 95% CI = [.06, .20], $d = 1.26$ (see Figure 4A). That is, participants in the tested group more frequently jointly recalled information learned from the same lecture segment than with information learned from other lecture segments. We also examined integration of information across different lecture segments by using participant responses from cued recall. As previously mentioned, integration was assessed by examining the amount of information that participants recalled that was not directly derived from the presented slide, but rather was mentioned in relation to other lecture slides. Within participants' responses, integration of material on the presented slide with material from other slides was often indicated by the phrase, "This is related to an earlier/later point the professor made about . . .", demonstrating an explicit link between two units of information learned across different lecture segments. Using this measure, we found that the tested group integrated significantly more units of information than the restudy group ($M_{\text{tested}} = 12.83$ units, $SE = 1.33$; $M_{\text{restudy}} = 4.67$ units, $SE = .97$), $t(34) = 4.97$, $p < .001$, 95% CI = [4.83, 11.51], $d = 1.66$ (see Figure 4B).

Across both portions of the final test, the tested group performed significantly better than the restudy group. However, we did not find a significant difference in mind wandering rates as assessed by random probes during the lecture. Next, we explored the relationship between mind wandering and final test performance in the tested and restudy groups.

Relationship between mind wandering and final test performance. In the tested group, we did not find a significant relationship between mind wandering and free recall performance, $r(16) = .095$, $p > .5$, 95% CI = [−.39, .54]. However, in the restudy group we observed a strong negative relationship between mind wandering and free recall performance, $r(16) = -.695$, $p = .001$, 95% CI = [−.88, −.34] (see Figure 5A). Using the Fisher's r-to-z transformation, we found that the correlation coefficients were significantly different from one another ($z = 2.61$, $p < .01$).

In the tested group, we also did not find a relationship between mind wandering and cued recall performance, $r(16) = -.139$, $p > .5$, 95% CI = [−.57, .35]. However, in the restudy group we once again observed a strong negative relationship between mind wandering and cued recall performance, $r(16) = -.70$, $p = .001$, 95% CI = [−.88, −.35] (see Figure 5B). Using the Fisher's r-to-z transformation, we found that the correlation coefficients were significantly different from one another ($z = 1.99$, $p < .05$).

These data suggest that mind wandering may have differential effects on final test performance in the tested and restudy groups. Whereas mind wandering did not seem to affect final test performance in the tested group, mind wandering was strongly correlated with final test performance in the restudy group, such that greater mind wandering rates were related to worse performance on a final test.

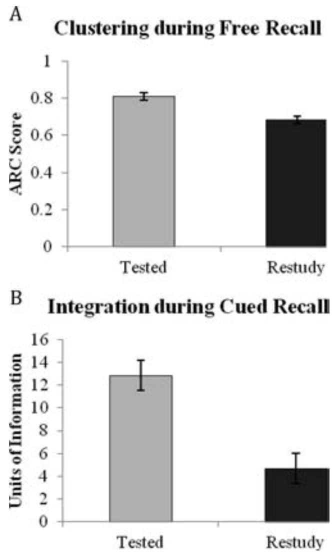

Figure 4. Integration of learned information in Experiment 1 as assessed by (Panel A) mean adjusted-ratio-of-clustering (ARC) scores of free recall output, where the range of ARC scores is 0 (no clustering) to 1 (perfect clustering) and (Panel B) mean units of information integrated from other slides during cued recall. Error bars represent one standard error of the mean.

Discussion

Overall, the results of Experiment 1 show that interpolated testing increased the amount of note-taking behaviors, boosted recall performance after the final lecture segment and during a final cumulative test, and improved the degree of clustering of related information learned within the same segment and integration of related information across different segments of the lecture. Furthermore, mind wandering differentially affected final test performance in the tested and restudy groups, such that mind wandering hurt performance in the restudy group more so than in the tested group.

In contrast to findings by Szpunar et al. (2013), we did not find significant differences in mind wandering rates between the tested and restudy groups. However, the mind wandering rates observed in Experiment 1 were very low (i.e., 20.8% in the tested group and 23.6% in the restudy group), compared to the mind wandering rates initially reported by Szpunar et al. (2013; i.e., 19% in tested group and 39% in restudy group). One notable difference between the two experiments that may contribute to this discrepancy is the use of different video-recorded lectures; whereas Szpunar et al. (2013) used a 20-min statistics lecture, in the current experiment we used a 40-min public health lecture. It has been shown that interest can affect rates of mind wandering during presentation of material (Giambra & Grodsky, 1989; Hollis & Was, 2016; Unsworth & McMillan, 2013). Thus, one possibility is that interpolated testing may have differential effects on mind wandering depending on the degree of participant interest in the lecture material. That is, interpolated testing may not benefit participants to the same extent when interest in the lecture is already high.

Alternatively, our assessment of "mind wandering" may be too broad. In Experiment 1, *mind wandering* was defined as instances where participants' attention strayed from the lecture content (i.e., if they were thinking about something other than what the lecturer was discussing during the probe's appearance). We ran a pilot study ($N = 36$) to assess the extent to which participants learning from a statistics lecture (same lecture used by Szpunar et al., 2013) mind wandered about lecture-unrelated content. Participants in a tested and restudy group both exhibited mind wandering content that was strongly associated with lecture-unrelated thoughts ($M_{\text{tested}} = 80\%$, $M_{\text{restudy}} = 83\%$). It is possible that use of a public health lecture may provide greater opportunities for variation in the content of "mind wandering" thoughts. For example, our pilot data indicate that participants can think about topics completely unrelated to the lecture, or simply zone out, both of which are examples of mind wandering defined as TUT (Smallwood et al., 2003). However, it is also possible that participants may engage in other thoughts that are related to the lecture. Notably, the content of different lecture segments was more interrelated in the public health lecture used in the current experiment than in the statistics lecture used by Szpunar et al. (2013), such that each segment frequently built upon the content of previous segments. Furthermore, public health is a topic that is highly relevant to everyday life. Thus, the content of the lecture is conducive to thinking back to earlier segments of the lecture and for relating the content of the lecture to one's own life, which are akin to TRT or TRI (Smallwood et al., 2003). As such, broadly probing "mind wandering"

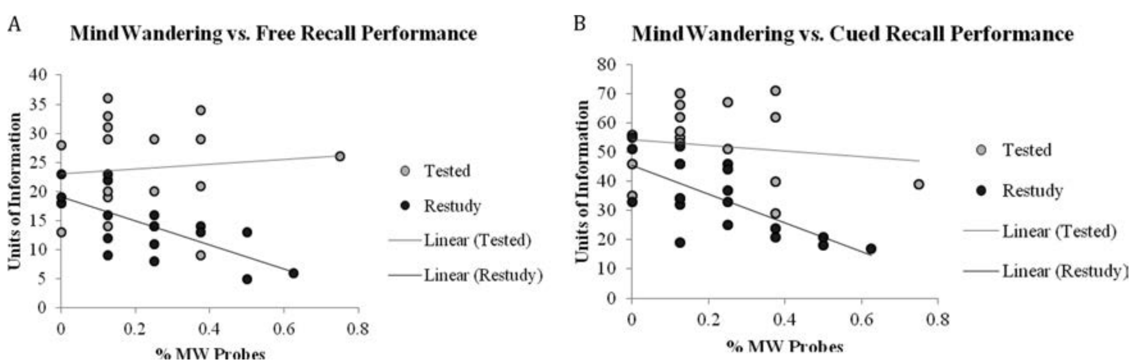

Figure 5. Experiment 1 correlations between mind wandering and mean units of information recalled during (Panel A) free recall and (Panel B) cued recall in tested (light gray) and restudy (dark gray) groups.

may not fully capture existing nuances in the types of thoughts that participants might have while watching the video-recorded lecture.

In Experiment 1, the final test was structured such that participants were first given an opportunity to freely recall information prior to elaborating upon individual lecture slides. However, this structure introduces a potential confound in our assessment of integration using cued recall. Specifically, freely recalling information for 7 min prior to completing cued recall may play a role in facilitating the integration of information from different parts of the lecture (i.e., participants may be clustering their recall across lecture segments during free recall). We can easily address this issue by reversing the order of the two final test components, which we implemented in Experiment 2.

Given that the tested and restudy groups did not significantly differ in their responses to mind wandering probes, why did we observe a difference in final test performance? First, the tested group took more notes than the restudy group; thus, one possibility is that differences in note-taking behaviors between the two groups could contribute to differences in final test performance. Alternatively, as discussed above, qualitative differences in the types of thoughts that participants engage in during the lecture might also explain the differential relationship between mind wandering and final test performance in the tested and restudy groups. Thus, in Experiment 2, we aimed to replicate the effects of interpolated testing on final test performance while more directly probing the content of participants' thoughts as they learn from a video-recorded lecture, and after changing the structure of the final test.

Experiment 2

In Experiment 2, we aimed to replicate and extend the results of Experiment 1 while more directly probing the content of participants' thoughts as they are engaged in watching the video-recorded lecture. Overall, the methods used in Experiment 2 are very similar to those of Experiment 1, with differences highlighted in the following text.

Method

Participants. A total of 39 healthy undergraduate students (ages 18 to 25, $M = 19.8$ years, 23 female) were recruited from Harvard College and Boston University. Participants were paid or received course credit for their participation. All participants had normal vision and no history of neurological or psychological impairment. In order to detect effects of interpolated testing on differences in thought probe frequency, we based our sample size on the same power analysis performed prior to Experiment 1 data collection (from Szpunar et al., 2013; $d = 1.05$, power $> .80$, $\alpha = .05$, two-tailed), which indicated that using a sample size of 16 participants per group would be sufficient. Data collection was stopped once it was determined that approximately enough usable participants had been run to reach this number. A total of 3 participants were excluded because of experimental glitches, leaving 36 participants in the final sample.

Design and procedure. The procedure for Experiment 2 was very similar to that of Experiment 1. The experimental session involved watching the same 40-min lecture used in Experiment 1 that was interspersed with thought probes and other activities, followed by a two-part final cumulative test.

Thought probes. Participants were informed that a visual cue indicated by the phrase "What are you thinking about right now?" would appear on the computer screen at random points during the lecture. Upon seeing the cue, they were asked to indicate their current thoughts by choosing one of five options: (1) what the professor is saying right now, (2) something related to an earlier part of the lecture, (3) relating the lecture to my own life, (4) something unrelated to the lecture, and (5) zoning out. Before beginning the experiment, the experimenter confirmed that participants understood the meaning of each probe option. The visual cue remained on the screen for 5 s as the lecture continued and was accompanied by an auditory cue (i.e., a bell) to ensure that students noticed the cue. Experiment 2 used the same four probe sequences from Experiment 1, and the presentation of thought probe sequences was counterbalanced across participants.

Final cumulative test. As in Experiment 1, the final cumulative test was administered in two parts. To avoid the possibility that freely recalling information for 7 min prior to completing cued recall may have facilitated the integration of information from various parts of the lecture, in Experiment 2 we reversed the order of the two portions of the final test. Participants first completed cued recall, where they were given 1 min to elaborate on 16 individually presented lecture slides. Afterward, participants completed 7 min of free recall. For both parts of the final test, participants typed information that was recalled in a text box provided on the computer screen. Participants' responses were scored by the same two raters from Experiment 1 (standardized Cronbach's $\alpha = .986$). Rater 1 scored 47% of participant responses, and rater 2 scored 53% of participant responses.

Results

Demographics. The tested and restudy groups did not differ in their levels of education, $t(26.74) = -.48$, $p > .5$, 95% CI = [−1.46, .91], $d = 0.16$, interest in the lecture $t(34) = -.39$, $p > .5$, 95% CI = [−1.52, 1.04], $d = 0.13$, or perceived difficulty in comprehending lecture material, $t(34) = -.17$, $p > .5$, 95% CI = [−1.15, .97], $d = 0.06$.

Thought probes and note taking. We first assessed mind wandering as defined in Experiment 1 (i.e., instances where participants were not directly focused on what the lecturer was discussing) by comparing the proportion of probe responses to choice 1 (i.e., thoughts about what the lecturer is saying right now) to the sum of the proportion of probe responses to choices 2 through 5 (i.e., thoughts related to an earlier part of the lecture, thoughts relating the lecture to my own life, thoughts unrelated to the lecture, and zoning out). As in Experiment 1, a series of two-tailed t tests revealed no significant difference in mind wandering between the tested and restudy groups ($M_{\text{tested}} = 22.9\%$, $SE = .04$; $M_{\text{restudy}} = 25\%$, $SE = .05$), $t(34) = -.33$, $p > .5$, 95% CI = [−.15, .11], $d = 0.11$ (see Figure 6A).

In segregating thought probes into five choices, we categorized choices 2 (i.e., thoughts about something related to an earlier part of the lecture) and 3 (i.e., thoughts relating the lecture to my own life) as lecture-related thoughts, because both types of thoughts require participants to pay attention to the lecturer's discussion to relate the content to other parts of the lecture or to aspects of their own lives. On the other hand, choices 4 (i.e., thoughts unrelated to the lecture) and 5 (i.e., zoning out) reflect *lecture-unrelated*

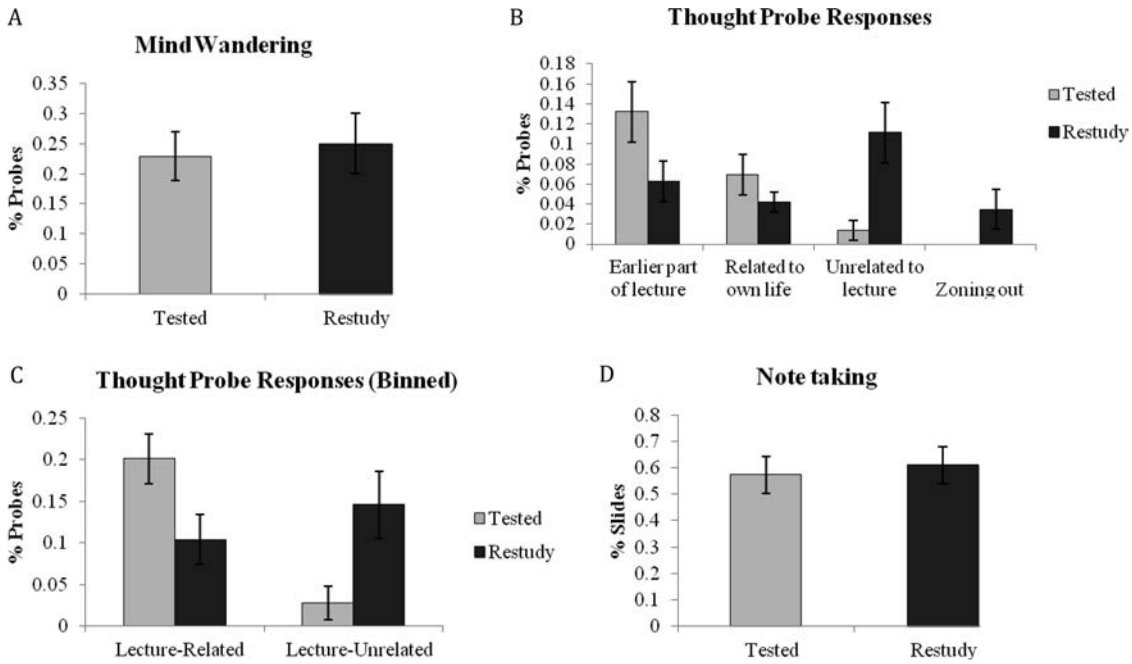

Figure 6. Experiment 2 mean proportions (out of 100%) of (Panel A) mind wandering as defined in Experiment 1 (i.e., sum of responses to thought probe choices 2 through 5); (Panel B) distribution of responses to thought probe choices 2 through 5; (Panel C) binned responses to thought probes that indicate lecture-related (i.e., probe choices 2 and 3) and lecture-unrelated (i.e., probe choices 4 and 5) thoughts; and (Panel D) note taking in tested and restudy groups. Error bars represent one standard error of the mean.

thoughts, or instances of mind wandering as typically defined in the literature. In order to assess group differences in the types of thoughts that deviated from what the lecturer was discussing at the time of the probe's appearance (i.e., probe choices 2 through 5), we first conducted a 2 (group: tested vs. restudy) × 4 (thought probes: choices 2, 3, 4, and 5) repeated-measures analysis of variance. We found a significant main effect of group, $F(3, 32) = 4.52$, $p < .01$, $\eta^2 = .30$, as well as a significant interaction of Group × Thought probes, $F(3, 32) = 4.64$, $p < .01$, $\eta^2 = .30$. Post hoc two-tailed *t* tests revealed that the tested group reported more thoughts related to an earlier part of the lecture ($M_{\text{tested}} = 13.2\%$, $SE = .03$) than the restudy group ($M_{\text{restudy}} = 6.3\%$, $SE = .02$), although this difference reached only trending significance, $t(28.37) = 2.01$, $p = .054$, 95% CI = [−.001, .14], $d = 0.67$. We also found that the restudy group ($M_{\text{restudy}} = 11.1\%$, $SE = .03$) reported significantly more thoughts that were unrelated to the lecture than the tested group ($M_{\text{tested}} = 1.4\%$, $SE = .01$), $t(20.03) = -2.93$, $p < .01$, 95% CI = [−.17, −.03], $d = 0.98$ (see Figure 6B). In categorizing responses to probes 2 and 3 as lecture-related thoughts and responses to probes 4 and 5 as lecture-unrelated thoughts, we found that the tested group ($M_{\text{tested}} = 20.14\%$, $SE = .03$) reported significantly more lecture-related thoughts than the restudy group ($M_{\text{restudy}} = 10.42\%$, $SE = .03$), $t(34) = 2.31$, $p < .05$, 95% CI = [.01, .18], $d = 0.77$. On the other hand, the restudy group ($M_{\text{restudy}} = 14.6\%$, $SE = .04$) reported significantly more lecture-unrelated thoughts than the tested group ($M_{\text{tested}} = 2.78\%$, $SE = .02$), $t(22.22) = -2.69$, $p < .05$, 95% CI = [−.21, −.03], $d = 0.90$ (see Figure 6C).

In contrast to Experiment 1, the tested and restudy groups did not differ significantly in the proportion of lecture slides on which they took notes ($M_{\text{tested}} = 57.4\%$, $SE = .07$; $M_{\text{restudy}} = 61.1\%$, $SE = .07$), $t(34) = -.39$, $p > .5$, 95% CI = [−.23, .16], $d = 0.13$ (see Figure 6D). Thus, observed differences in final test performance cannot be attributed to differential note-taking behaviors.

Eighth segment recall. Next, we compared initial recall of information after the eighth segment of the lecture between the tested and restudy groups. As in Experiment 1, the tested group recalled significantly more units of information than the restudy group ($M_{\text{tested}} = 7.67$ units, $SE = .54$; $M_{\text{restudy}} = 4.50$ units, $SE = .40$), $t(34) = 4.71$, $p < .001$, 95% CI = [1.80, 4.53], $d = 1.57$ (see Figure 7A).

Final test. We also assessed performance on the two portions of the final cumulative test in the tested and restudy groups. As previously mentioned, in Experiment 2 we reversed the order such that participants first completed cued recall and subsequently completed free recall. Overall, the tested group recalled significantly more units of information than the restudy group during free recall ($M_{\text{tested}} = 20.94$ units, $SE = 1.55$; $M_{\text{restudy}} = 13.58$ units, $SE = 1.31$), $t(34) = 3.63$, $p = .001$, 95% CI = [3.23, 11.49], $d = 1.21$ (see Figure 7B). The tested group also recalled significantly more units of information than the restudy group during cued recall ($M_{\text{tested}} = 45.11$ units, $SE = 2.79$; $M_{\text{restudy}} = 30.00$ units, $SE = 2.70$), $t(34) = 3.90$, $p < .001$, 95% CI = [7.23, 22.99], $d = 1.30$ (see Figure 7C).

As in Experiment 1, we further assessed whether improved learning during the eighth lecture segment (as assessed by eighth segment recall) persisted over a delay by examining recall of information during free recall that originated from the eighth segment of the lecture. The tested group recalled significantly more units of information from the eighth lecture segment than the

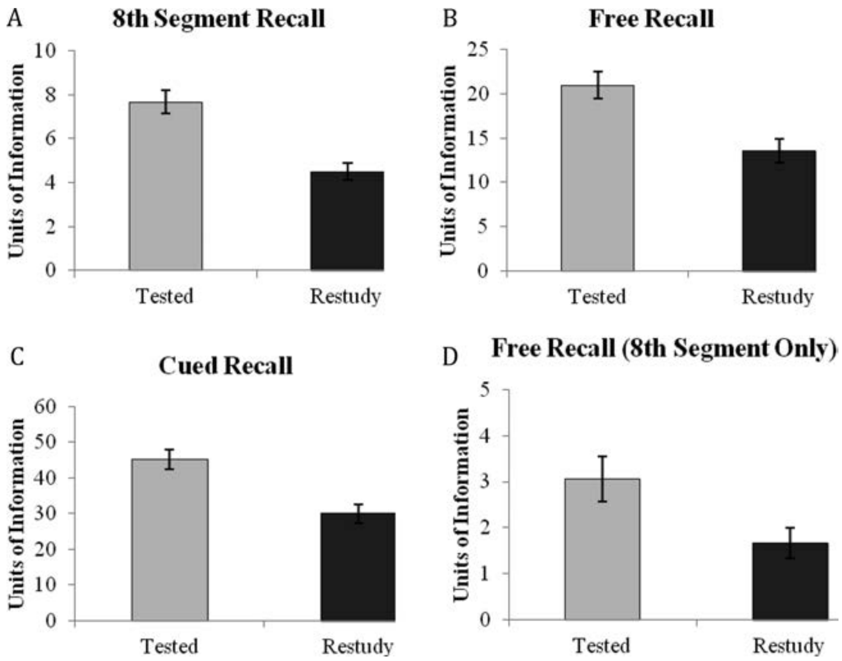

Figure 7. Experiment 2 mean units of information recalled: (Panel A) after the eighth segment of lecture, (Panel B) during free recall, (Panel C) during cued recall, and (Panel D) during free recall that originated from the eighth segment of the lecture. Error bars represent one standard error of the mean.

restudy group during free recall (M_{tested} = 3.06 units, *SE* = .49; M_{restudy} = 1.67 units, *SE* = .32), $t(34) = 2.35$, $p < .05$, 95% CI = [0.19, 2.59], $d = 0.78$ (see Figure 7D). Thus, the benefits of testing on free recall persisted across a short delay.

Integration. During free recall, we found that participants in the tested group (M_{ARC} = .73, *SE* = .03) were more likely than were participants in the restudy group (M_{ARC} = .61, *SE* = .02) to cluster their output such that they frequently jointly recalled information from the same segment, $t(34) = 3.65$, p = .001, 95% CI = [.06, .19], $d = 1.22$ (see Figure 8A). During cued recall, we also found that the tested group integrated significantly more units of information than the restudy group (M_{tested} = 8.44 units, *SE* = 1.28; M_{restudy} = 3.00 units, *SE* = .71) by relating an individually presented slide to other slides in the lecture, $t(34) = 3.72$, $p < .001$, 95% CI = [2.47, 8.42], d = 1.24 (see Figure 8B).

As in Experiment 1, the tested group performed significantly better and integrated more information than did the restudy group on both portions of the final test. Given the differences in the types of thoughts that participants reported during the lecture, next we explored the relationship between different types of thoughts (i.e., lecture-related vs. lecture-unrelated) and final test performance in the tested and restudy groups.

Relationship between types of thought and final test performance. Overall, we found a strong negative relationship between lecture-unrelated thoughts (i.e., responses to probe

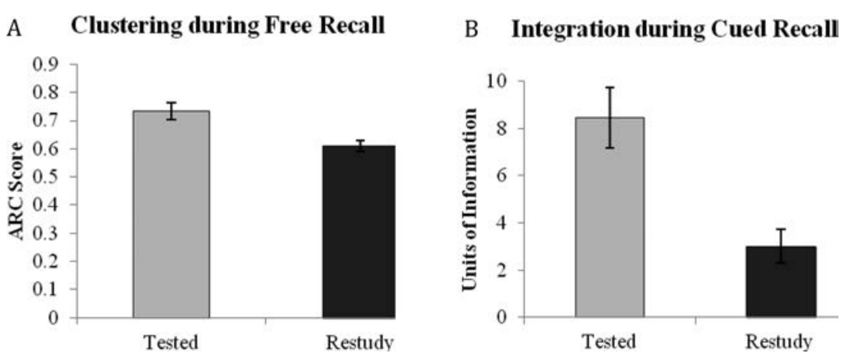

Figure 8. Integration of learned information in Experiment 2 as assessed by (Panel A) mean adjusted-ratio-of-clustering (ARC) scores of free recall output, where the range of ARC scores is 0 (no clustering) to 1 (perfect clustering) and (Panel B) mean units of information integrated from other slides during cued recall. Error bars represent one standard error of the mean.

choices 4 and 5) and performance on both free recall, $r(34) = -.541$, $p = .001$, 95% CI = [−.74, −.26] (see Figure 9A), and cued recall, $r(34) = -.595$, $p < .001$, 95% CI = [−.77, −.33] (see Figure 9B). It is important to note that among the participants who reported lecture-related thoughts (i.e., responses to probe choices 2 and 3; $N = 27$), we observed a positive relationship between lecture-related thoughts and performance on free recall, $r(25) = .45$, $p < .05$, 95% CI = [.09, .71] (see Figure 10A), and cued recall, $r(25) = .458$, $p < .05$, 95% CI = [.10, .71] (see Figure 10B). We also found a positive relationship between lecture-related thoughts and the amount of integrated information during cued recall, $r(25) = .47$, $p < .05$, 95% CI = [.11, .72] (see Figure 10C).

These data suggest that lecture-unrelated thoughts may predominantly drive the negative consequences of mind wandering on final test performance. However, lecture-related thoughts may in fact be beneficial for learning.

Discussion

The results of Experiment 2 replicate and extend the results of Experiment 1 after segregating thought probes into five choices to more clearly assess participants' thoughts as they learn from the lecture. First, the tested group as compared to the restudy group recalled significantly more units of information after the eighth segment of the lecture and during both components of the final test, and also more frequently clustered their responses during free recall. Critically, the tested group integrated significantly more information during cued recall than the restudy group, even after reversing the order of the final test components. Thus, as in Experiment 1, we show that interpolated testing can boost performance on a later assessment of learning during the lecture. In contrast to Experiment 1, here we did not find a significant difference in the amount of notes taken between the tested and restudy groups. Thus, differences in final test performance cannot be attributed to differences in the quantity of notes taken while participants viewed the lecture video.

Second, although both groups were reportedly focused on the lecturer's discussion for equal amounts of time, the tested group reported more thoughts related to an earlier part of the lecture, whereas the restudy group reported more thoughts unrelated to the lecture. Furthermore, whereas lecture-unrelated thoughts (i.e., thoughts about an earlier part of the lecture and relating the lecture to one's own life) hurt performance on the final test, lecture-related thoughts (i.e., thoughts unrelated to the lecture and instances of zoning out) were positively related to final test performance. Thus, not all thoughts that participants engage in while watching a video-recorded lecture may be detrimental to learning.

General Discussion

Across two experiments, we demonstrate that interpolated testing can boost learning during a video-recorded lecture relative to restudying the same material, as assessed by the amount of information recalled after the eighth segment of the lecture and during a final cumulative test. The finding that testing enhances learning replicates a large body of literature on the testing effect, which is the idea that repeated retrieval or testing during learning enhances long-term retention of material (for review, see Roediger & Karpicke, 2006).

However, of particular interest is the finding that interpolated testing can boost integration of information. First, during free recall, participants in the tested group were more likely than participants in the restudy group to cluster their output by lecture segment. That is, tested participants more often followed a piece of information recalled from one lecture segment with another related item recalled from the same segment, as opposed to information from a different lecture segment. These results are consistent with existing literature showing that testing can enhance the discriminability of word lists during recall, such that participants are better able to segregate their recall of words that originated from different lists (e.g., Szpunar et al., 2008). In the context of our current experiments, increased clustering of free recall responses by lecture segment may indicate greater integration of related information that was learned within each individual lecture segment. However, this finding does not preclude the possibility that interpolated testing may also boost integration of information across lecture segments. To address this possibility, we used a test that more directly requested participants to integrate information across lecture segments. During cued recall, the tested group more readily recalled and integrated relevant information from other slides when elaborating upon an individually presented slide. Existing

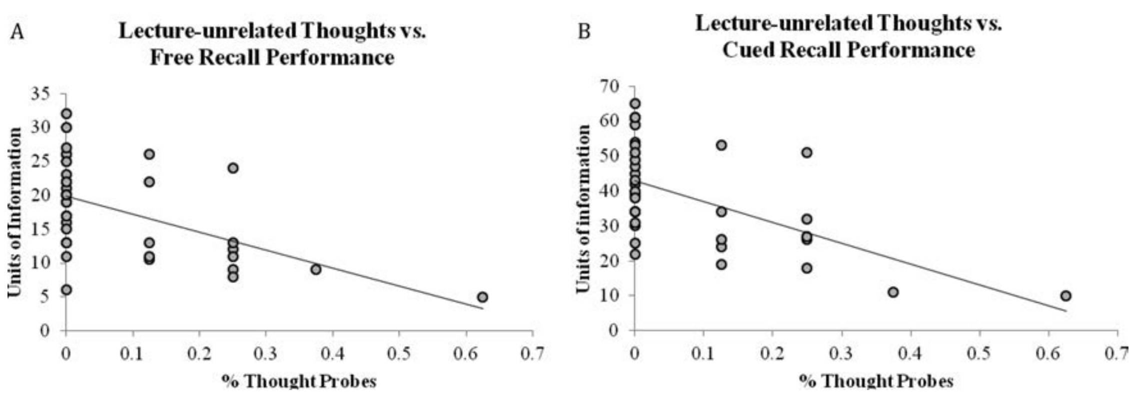

Figure 9. Experiment 2 correlations between lecture-unrelated thoughts (i.e., responses to thought probe choices 4 and 5) and mean units of information recalled during (Panel A) free recall and (Panel B) cued recall.

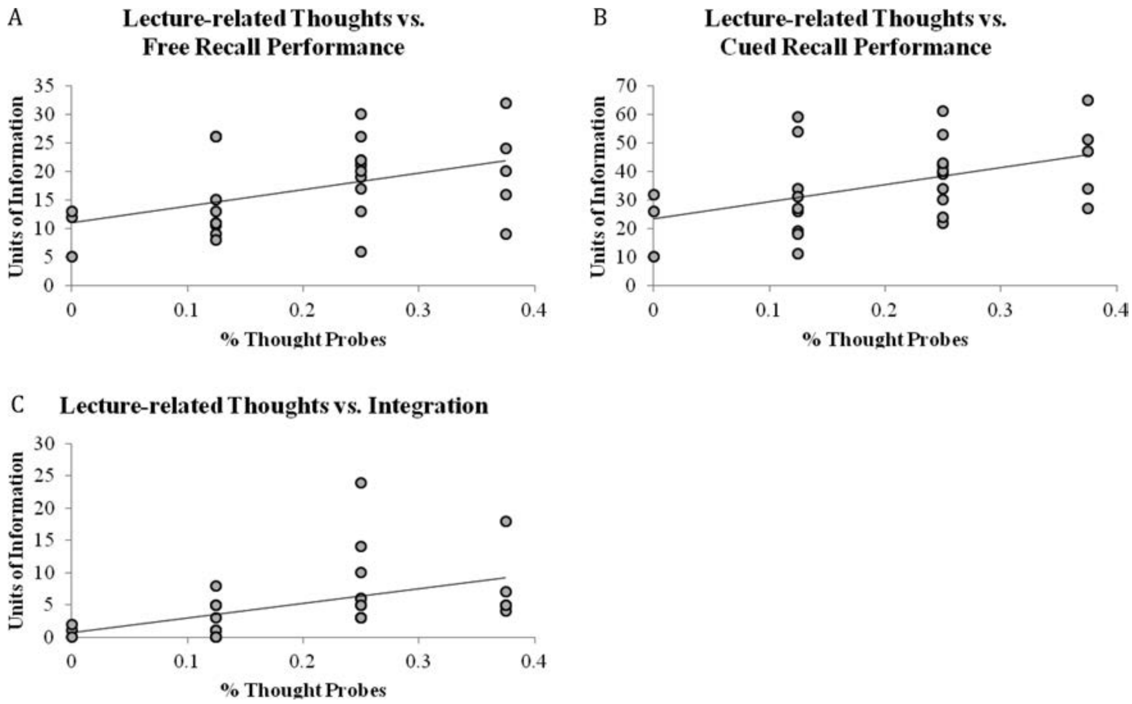

Figure 10. Experiment 2 correlations between lecture-related thoughts (i.e., responses to thought probe choices 2 and 3) and mean units of information recalled during (Panel A) free recall, (Panel B) cued recall, and (Panel C) information integrated from other slides during cued recall.

research has shown that interpolated testing during the presentation of consecutive word lists can enhance the extent to which participants cluster their recall of related words that were presented across different word lists on a final cumulative test (Zaromb & Roediger, 2010; for related discussion, see Wahlheim, 2015) and possibly enhance comprehension of prose presented in an interpolated fashion (Wissman et al., 2011; Wissman & Rawson, 2015; see Experiments 6 and 7). Here, we demonstrate for the first time that interpolated testing can also improve the integration of related information that is learned across different segments within a video-recorded lecture. One possible mechanism by which this may occur is that regularly retrieving information after each lecture segment serves to reactivate previously learned information, allowing participants to hold the information in mind as they learn from each new lecture segment to a greater degree than restudying the material (cf. Wahlheim, 2015). As a result, participants who were tested may have been more effective at noticing the relations between related concepts across different lecture segments, thus promoting the linkage of newly learned information to material that was learned in a previous lecture segment. In this manner, interpolated testing may facilitate the integration of related concepts during learning and during the final test.

Note that this finding directly contrasts with other existing hypotheses regarding testing and relational processing, mainly that relative to restudying material, testing may in fact impair integration because testing tends to direct one's attention toward item-specific processing that may detract from noticing the relations between items (e.g., Peterson & Mulligan, 2013), or because testing in an interpolated fashion may disrupt the individual from connecting ideas from different sections (e.g., Wissman & Rawson, 2015). However, experiments directly testing these hypotheses have not found definitive supporting evidence (e.g., Wissman & Rawson, 2015), or have failed to replicate previous findings (e.g., Rawson et al., 2015). Given the inconsistent results and large differences in experimental materials used across these various studies, more research still needs to be done on examining the effect of interpolated testing on information integration in more complex learning materials.

Another notable finding in the current experiments concerns the types of thoughts that participants engaged in as they watched the lecture video. First, in Experiment 1 we found similar rates of mind wandering (i.e., defined as instances in which participants were not directly focused on the lecturer's discussion when the mind wandering probe appeared) in the tested and restudy groups. However, whereas mind wandering was negatively related to final test performance in the restudy group, mind wandering did not seem to significantly impact final test performance in the tested group. After segregating thought probes into five choices in Experiment 2, we observed qualitative differences in the types of thoughts that participants reported while watching the lecture. Specifically, the tested group reported more instances of lecture-related thoughts (i.e., thoughts related to an earlier part of the lecture and thoughts relating the lecture to one's own life), whereas the restudy group reported more instances of lecture-unrelated thoughts (i.e., thoughts unrelated to the lecture and zoning out). Lecture-unrelated thoughts correspond to standard TUTs, as defined in the literature, and are indicative of mind wandering (e.g., Smallwood et al., 2003); it is thus unsurprising that engaging in

lecture-related thoughts was negatively related to final test performance, an observation consistent with prior research showing that mind wandering is associated with poorer learning outcomes (Hollis & Was, 2016; Lindquist & McLean, 2011; Risko et al., 2012). However, lecture-related thoughts may share more similarities with either TRT or TRI (Smallwood et al., 2003). The observation of lecture-related thoughts in the tested group may reflect the enhanced accessibility of prior material facilitated by testing, and notably, engagement in lecture-related thoughts was positively related to final test performance. Overall, these results demonstrate for the first time that not all thoughts that participants engage in while watching a video-recorded lecture may be detrimental to learning, and further research is needed to elucidate how different types of thoughts (both related and unrelated to the task) might enhance or interfere with learning.

In two studies, we show that interpolated testing can benefit learning from a video-recorded lecture in several ways, including influencing the types of thoughts that participants engage in during the lecture, improving learning of lecture material, and enhancing integration of material both within a single lecture segment and across different lecture segments. Moving forward, future work is needed to address the effect of interpolated testing on learning and integration at a longer delay, the role of feedback, and the transfer of knowledge learned within a lecture to answer related questions in different contexts. Furthermore, it is an open question whether interpolated testing would have similar effects when using lecture materials other than statistics and public health. Nonetheless, we have now demonstrated that interpolated testing is generally associated with focusing attention and boosting learning and integration of study materials. With the advent of many online learning platforms and the use of video-recorded lectures to supplement classroom lectures, these results may inform the improvement of the quality of learning in both classroom and online learning contexts alike.

References

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into EdX's first MOOC. *Research and Practice in Assessment, 8,* 13–25.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory, 18,* 385–393. http://dx.doi.org/10.1080/09658211003702163

Bunce, D. M., Flens, E. A., & Neiles, K. Y. (2010). How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education, 87,* 1438–1443. http://dx.doi.org/10.1021/ed100409p

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. http://dx.doi.org/10.1037/0278-7393.33.2.431

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*. Advance online publication. Retrieved from http://www.amstat.org/publications/jse/v2n2/gal.html

Giambra, L. M. (1995). A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness and Cognition, 4,* 1–21. http://dx.doi.org/10.1006/ccog.1995.1001

Giambra, L. M., & Grodsky, A. (1989). Task-unrelated images and thoughts while reading. In J. Shorr, P. Robin, J. A. Connek, & M. Wolpin (Eds.), *Imagery: Current perspectives* (pp. 26–31). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-0876-6_3

Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning and Instruction, 42,* 104–112. http://dx.doi.org/10.1016/j.learninstruc.2016.01.007

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science, 22,* 10–15. http://dx.doi.org/10.1177/0963721412463228

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497–514. http://dx.doi.org/10.1016/S0022-5371(81)90138-9

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421–445. http://dx.doi.org/10.1006/jmla.1993.1023

Hunt, R. R., & Rawson, K. A. (2011). Knowledge affords distinctive processing in memory. *Journal of Memory and Language, 65,* 390–405. http://dx.doi.org/10.1016/j.jml.2011.07.001

Johnstone, A. H., & Percival, F. (1976). Attention breaks in lectures. *Education in Chemistry, 13,* 49–50.

Khan, S. (2012). *The one world school house: Education reimagined*. London, UK: Hodder & Stoughton.

Koller, D. (2011). *Death knell for the lecture: Technology as a passport to personalized education*. http://www.nytimes.com/2011/12/06/science/daphne-koller-technology-as-a-passport-to-personalized-education.html?pagewanted=all&_r=1

Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences, 21,* 158–167. http://dx.doi.org/10.1016/j.lindif.2010.12.006

Masson, M. E. J., & McDaniel, M. A. (1981). The role of organizational processes in long-term retention. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 100–110. http://dx.doi.org/10.1037/0278-7393.7.2.100

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103,* 399–414. http://dx.doi.org/10.1037/a0021782

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136,* 188–197. http://dx.doi.org/10.1037/a0018298

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

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. http://dx.doi.org/10.1037/a0021801

Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1287–1293. http://dx.doi.org/10.1037/a0031337

Rawson, K. A., Wissman, K. T., & Vaughn, K. E. (2015). Does testing impair relational processing? Failed attempts to replicate the negative testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1326–1336. http://dx.doi.org/10.1037/xlm0000127

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology, 26,* 234–242. http://dx.doi.org/10.1002/acp.1814

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. http://dx.doi.org/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. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

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

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology, 1,* 60–71. http://dx.doi.org/10.1037/stl0000011

Seli, P., Carriere, J. S. A., & Smilek, D. (2015). Not all mind wandering is created equal: Dissociating deliberate from spontaneous mind wandering. *Psychological Research, 79,* 750–758. http://dx.doi.org/10.1007/s00426-014-0617-x

Seli, P., Cheyne, J. A., Xu, M., Purdon, C., & Smilek, D. (2015). Motivation, intentionality, and mind wandering: Implications for assessments of task-unrelated thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1417–1425. http://dx.doi.org/10.1037/xlm0000116

Smallwood, J. (2010). Why the global availability of mind wandering necessitates resource competition: Reply to McVay and Kane (2010). *Psychological Bulletin, 136,* 202–207. http://dx.doi.org/10.1037/a0018673

Smallwood, J. M., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition, 12,* 452–484. http://dx.doi.org/10.1016/S1053-8100(03)00018-7

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13,* 657–690. http://dx.doi.org/10.1016/j.concog.2004.06.003

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132,* 946–958. http://dx.doi.org/10.1037/0033-2909.132.6.946

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications for online education. *Journal of Applied Research in Memory & Cognition, 3,* 161–164. http://dx.doi.org/10.1016/j.jarmac.2014.02.001

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 6313–6317. http://dx.doi.org/10.1073/pnas.1221764110

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. http://dx.doi.org/10.1037/a0013082

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 832–842. http://dx.doi.org/10.1037/a0029669

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43,* 27–38. http://dx.doi.org/10.3758/s13421-014-0455-5

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41,* 1–15. http://dx.doi.org/10.3758/s13421-012-0246-9

Wilson, K., & Korn, J. H. (2007). Attention during lectures: Beyond ten minutes. *Teaching of Psychology, 34,* 85–89. http://dx.doi.org/10.1177/009862830703400202

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. http://dx.doi.org/10.1037/xlm0000047

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. http://dx.doi.org/10.3758/s13423-011-0140-7

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

Received December 16, 2015
Revision received February 26, 2016
Accepted April 3, 2016 ■