Optimizing the Use of Interpolated Tests: The Influence of Interpolated Test Lag

Eric Lavigne and Evan F. Risko
University of Waterloo

The use of recorded lectures is increasing rapidly due to growth in online learning. One technique that can be used to improve learning from lectures is interpolated testing—the presentation of tests throughout a lecture. In the present investigation, we examine a critical question with respect to the implementation of interpolated testing. Where should the tests be located relative to the tested material? Specifically, we examine the influence of the lag between the presentation of the material in the lecture and the interpolated test. Across 2 experiments, we compare an immediate test condition (i.e., a test immediately after the relevant information is presented) and a delayed test condition (i.e., a test presented about 3 min after the relevant information is presented). When no feedback was provided, immediate interpolated testing was superior to delayed interpolated testing. There was no difference when feedback was provided. Implications of the present results for implementing interpolated testing in educational contexts are discussed.

Keywords: testing, lectures, memory

There has been massive growth in online learning in postsecondary education (e.g., Allen & Seaman, 2014). One of the primary pedagogical devices in the online learning domain is a recorded lecture. These lectures present material typically created by an instructor and instructional designer and made available via the Internet. Critically, recorded lectures provide opportunities for implementing teaching strategies that could enhance learning. One such strategy is the use of interpolated tests—tests that are inserted throughout the recorded lecture (Szpunar, Jing, & Schacter, 2014; Szpunar, Khan, & Schacter, 2013). Much research has demonstrated that testing in general can be used to enhance retention (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Butler, 2011; Roediger & Karpicke, 2006) and Szpunar and colleagues demonstrated this specifically in the context of interpolated tests. However, interpolated tests could be implemented in a number of different ways within a recorded lecture, thus raising questions regarding how best to use interpolated tests. We turn to this question in the present investigation. In particular we examine one of the first implementational questions one might encounter: Where in the lecture should the interpolated tests be located with respect to the study material?

In the first examination of interpolated testing in the context of lectures, Szpunar et al. (2013) conducted two experiments that involved participants learning statistics from a video lecture (viewed in a laboratory). The lecture was divided into four segments and participants were either given a test at the end of each segment (interpolated test condition) or an arbitrary arithmetic problem to solve (non-test condition). In the second experiment, they also added a restudy condition. They found that interpo-
lated testing resulted in better performance on the final test than in the non-testing and restudy conditions. They also found that interpolated testing reduced mind wandering and increased note taking frequency. Szpunar et al. (2014), using a recorded lecture (viewed in a laboratory), replicated and extended this work by demonstrating that interpolated testing helped reduce overconfidence. In both of these studies multiple questions were presented following a lecture segment. As such, the lag between information presentation and the testing of that information was variable. This need not be the case. For example, a test question could be presented immediately after the information is presented or systematically delayed by some amount of time. While no research has directly assessed which method is preferable in terms of retention, two influential ideas lead to two different predictions and relevant research has yielded an unclear picture.

There appears to be a general consensus around the idea that delayed testing benefits retention (Cepeda et al., 2009; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Rawson, Vaughn, & Carpenter, 2015). Delayed testing causes an increase in the difficulty of the test, which in turn results in better long-term retention (i.e., a desirable difficulty). In an extension of the Elaborative Retrieval Hypothesis (Carpenter, 2009, 2011), a popular account of the testing effect, Rawson, Vaughn, and Carpenter (2015) suggested that longer lags provide a greater opportunity for elaborative retrieval (i.e., activation of related information) than shorter lags. From this perspective, delayed interpolated tests within a lecture could be expected to be superior to immediate interpolated tests. Alternatively, research on testing has also revealed that retrieval success represents an important factor in successful retention (e.g., Pyc & Rawson, 2009) and immediate testing is likely to lead to greater retrieval success. The influence of retrieval success can also be interpreted within the elaborative retrieval framework where low retrieval success contexts would fail to generate elaborative retrieval (Rawson et al., 2015). From this perspective, immediate interpolated tests within a lecture would be expected to be superior to delayed interpolated tests.

In research examining expanding retrieval practice, Karpicke and Roediger (2007) asked participants to learn vocabulary word pairs and each pair was presented according to one of several different schedules. Two of these schedules involved single tests (one immediate and one delayed), which to some extent mimics the contrast between immediate and delayed interpolated testing, and in almost all cases a delayed test led to superior retention of information from the lecture. Karpicke and Roediger (2010) examined a similar question using memory for expository texts. In their second experiment, they had participants study nine passages, using various testing schedules and manipulating the presence of feedback. Unlike Karpicke and Roediger (2007), in the single test conditions (one immediate test or one delayed test), overall there was no difference between the immediate and delayed single test condition. However, there was an interaction with lag that appeared to reflect a small benefit for immediate testing with no feedback and a larger benefit for delayed testing when feedback was provided. Recent work has also compared retention as a function of whether testing occurred interpolated throughout a study session (i.e., akin to immediate testing) or delayed until the end of the study session (Healy, Jones, Lalchandani, & Tack, 2017; Weinstein, Nunes, & Karpicke, 2016; Wissman & Rawson, 2015). For example, Weinstein, Nunes, and Karpicke (2016) compared testing administered after each relevant fact in a slideshow with testing administered at the end of the slideshow. Healy, Jones, Lalchandani, and Tack (2017) and Wissman and Rawson (2015) followed similar procedures. Across these studies, performance on final retention tests either showed no difference between conditions or a benefit for interpolated testing over end-of-session testing. Thus, taken together, research on expanding retrieval practice (Karpicke & Roediger, 2007, 2010) and interpolated versus end-of-session testing (Healy et al., 2017; Weinstein et al., 2016; Wissman & Rawson, 2015), which are arguably similar to the contrast between immediate and delayed interpolated testing, have yielded a somewhat unclear picture with respect to whether immediate versus delayed interpolated tests would yield any differential benefits and in which direction.

Present Investigation

None of the research reviewed above compared immediate versus delayed interpolated testing directly. In the experiments reported here, participants were asked to study a recorded lecture online and were tested on the material afterward. In
Experiment 1, there were three between-subjects conditions represented by the presence or absence of interpolated tests and the delay between the presentation of information in the lecture and the interpolated test on that information. The no interpolated testing condition was included as a control to assess the benefits of interpolated testing. In the interpolated testing conditions, participants were presented with short answer questions either immediately after the information was presented in the lecture or following a 3-min delay. In addition, Experiment 1 provided participants with no feedback, whereas in Experiment 2 feedback was presented. Providing feedback is typical in classroom settings and, as it provides a re-exposure to the content, provided us with the opportunity to investigate whether re-exposure would modulate the influence of interpolated test lag. For example, immediate testing is likely to lead to a higher rate of successful recall during the lecture and thus a higher likelihood of re-exposure. Providing feedback could reduce any differential benefit of re-exposure (i.e., both conditions are re-exposed to the answer after the question). Thus, comparing feedback and no feedback conditions provide both practical value (i.e., how do the benefits/costs of immediate vs. delayed interpolated tests vary as a function of pedagogical strategy) and some insight into the potential mechanism underlying any differences between immediate and delayed testing.

Based on previous research, interpolated testing should lead to improved performance on the final test (relative to no interpolated tests). In addition, performance on the interpolated tests is expected to be greater in the immediate condition than in the delayed condition and performance on the final test should be greater in Experiment 2, where interpolated test feedback is provided, than Experiment 1 where no feedback is provided. The critical question is whether immediate testing or delayed testing yield better performance in terms of final test performance and whether this difference is modulated by the provision of feedback.

**Experiment 1**

**Method**

**Subjects.** We set out to collect 150 participants in the interpolated testing conditions in order to have sufficient power to detect a medium-sized effect ($d = 0.5$, power = 0.8). In Experiment 1, 250 (109 immediate; 108 delayed; 33 no testing control) individuals participated in exchange for $10.00. All participants took part online through Amazon Mechanical Turk. Participants varied in age, gender, ethnicity, and educational background.

**Design.** A 3 (Test Type: no interpolated test vs. immediate interpolated test vs. delayed interpolated test) between-subjects design was used.

**Stimuli.** The recorded lecture was administered using a software tool developed for this purpose. The video lecture that was used was the second lecture of an Introduction to Biology course taught from Massachusetts Institute of Technology’s MIT Open Courseware website (http://ocw.mit.edu/courses/biology/7-012-introduction-to-biology-fall-2004/video-lectures/lecture-2-biochemistry-1/). The topic of this particular lecture is Biochemistry. The first 23 min was shown to participants. A total of 16 short-answer questions were created based on the material in the recorded lecture for use as interpolated tests. Presentation of the question involved pausing the lecture and the question appearing overlaid on the lecture video. Participants could not interact or see the video until an answer was submitted. Video scrubbing was also disabled so participants could not rewatch or skip parts of the lecture. The final retention test consisted of the same 16 questions as those used for the interpolated tests but were presented in a multiple-choice format instead of the original short answer format. In order to establish a baseline of performance on this final test, we had 341 participants from the same participant population complete the final retention test with no lecture or interpolated questions. The mean (correct) test score was 41.5% CI [35.5%, 47.6%].

**Procedure.** The experiment was conducted online using Amazon’s Mechanical Turk system. After accepting our posted task, participants were given a brief overview of the study, as well as a link to a website. When they reached this website, participants were then given the detailed instructions on how to complete the recorded lecture portion of the study.

---

1 A total of 40 participants were collected in this no lecture baseline data, but six were excluded from the analysis because of failure to complete the final test, or duplicate submission.
They were told that questions may come up throughout the lecture and that they needed to answer these questions to the best of their ability. Participants were then instructed that they were to watch the lecture. Before doing so, they were asked to complete a demographic questionnaire. The demographic questionnaire asked participants to fill out information regarding their familiarity with the general subject of biochemistry (“How familiar were you with the material on chemistry and microbiology [e.g., molecule formations, bonds, thermodynamics?”] and how many school courses they have taken in both chemistry and biology (“Roughly how many biology/chemistry courses have you taken?”). Participants then watched the lecture and, depending on condition, either answered interpolated questions throughout (immediate and delayed interpolated testing) or not (control). In the immediate condition, the question appeared directly after the information was presented in the lecture. In the delayed condition, the question appeared approximately 3 min after the information was presented. For those in the control condition, the video played through without any pausing. After the lecture was finished, participants were taken to the final retention test. There was no time limit on the interpolated and final test questions. After completing the final test, participants were debriefed.

Results

In Experiment 1, 57 participants were excluded due to technical reasons (e.g., duplicate participation). After these exclusions, there were 88 participants in the immediate test type condition, 74 in the delayed test type condition, and 30 in the control condition. Interpolated tests were graded by two markers, one of whom graded all the answers while the other graded 25% of the answers. The raw percentage agreement (on answers graded by both markers) was 92%, with Kappa of 0.85. When a Levene’s test revealed a violation of the homogeneity of variance in the ANOVAs reported below, a correction was applied and did not alter the results qualitatively. The original results are reported. There was no difference across conditions in familiarity with the general subject of biochemistry (eight participants did not complete this information), $F(2, 181) = 1.0, p = .35, \eta^2_G = 0.01$, or how many school courses they had taken in chemistry, $F(2, 189) = 0.56$, $p = .57, \eta^2_G = 0.01$; or biology, $F(2, 189) = 1.38, p = .25, \eta^2_G = 0.01$. Data are presented in Figure 1.

**Interpolated testing scores.** A one-way between subject ANOVA (immediate vs. delayed) was conducted on interpolated test scores. There was a significant effect of interpolated test delay, $F(1, 160) = 185.43, p < .001, \eta^2_G = 0.54, d = 2.14$, such that participants in the immediate condition ($M = 63.9\%$, CI [61.3%, 66.4%]) scored significantly higher than those in the delayed condition ($M = 34.0\%$, CI [30.3%, 37.7%]).

**Final retention scores.** A one-way between subject ANOVA (Test Type: immediate vs. delayed vs. control) was conducted on final retention test scores. There was a significant effect of test type, $F(2, 189) = 13.62, p < .001, \eta^2_G = 0.13$, such that participants in the immediate interpolated test condition ($M = 75.0\%$, CI [71.6%, 78.4%]) performed better on the final test than those in the delayed interpolated test condition ($M = 67.8\%$, CI [63.8%, 71.8%]), $t(160) = 2.74, p = .007, d = 0.43$. Participants in the control condition ($M = 56.7\%$, CI [49.7%, 63.7%]) performed worse than individuals in the immediate test condition, $t(116) = 5.21, p < .001, d = 1.06$, and delayed test condition, $t(102) = 2.90, p = .005, d = 0.62$

**Question response time.** While not the focus of the present research, we also collected response times to each interpolated and final test question. We report an exploratory analysis here. With respect to the interpolated test ques-

![Figure 1](image-url)
effect of interpolated test delay, test response times. There was no significant (immediate vs. delayed) was conducted on final
the data. A one-way between-subjects ANOVA
ducted and resulted in the removal of 3.9% of
collected. The same outlier analysis was con-
removed because the response time data was not
response times greater than 2.5 standard devi-
tions from the participant mean, resulting in
of the remaining participants, 1.1% of ob-
servations were removed for the same reason.
An outlier analysis was conducted that excluded
exclusively for the personal use of the individual user and is not to be disseminated broadly.

Discussion

The results of Experiment 1 demonstrate that
in the context of a recorded lecture with no
feedback, immediate interpolated tests lead to
greater gains in retention than a delayed (3 min)
interpolated test. Both forms of interpolated
tests outperformed a no-test control, demonstrat-
ing again that interpolated testing in the
context of recorded lectures can benefit reten-
tion. A detailed discussion will follow Experi-
ment 2.

In Experiment 2, we wanted to test one poten-
tial explanation of the benefit of immediate
over delayed interpolated testing on the final
test. Specifically, in the immediate condition,
the associated benefit on the final test could
reflect the higher likelihood of being re-exposed
to the answer to the questions during the inter-
polated testing part of the experiment provided
the large difference in successful retrieval
across conditions. To test this account, we rep-
licated the immediate and delayed interpolated
conditions from Experiment 1 but added feed-
back. If the benefit of immediate interpolated
testing is a product of the increased likelihood
of re-exposure, then the advantage of immediate
over delayed interpolated testing should be re-
duced or eliminated when feedback is provided.
In this case, feedback can be conceptualized as
reducing the putative difference between the
immediate and delayed conditions in terms of
re-exposure afforded by the difference in suc-
cessful retrieval across conditions (i.e., all par-
ticipants are re-exposed to the answers after
each interpolated test). From a practical per-
spective, different instructional contexts may
make the provision of feedback more or less
desirable and, as such, understanding how this
variable modulates the influence of interpolated
test lag is valuable.

Experiment 2

Method

Subjects. In Experiment 2, 193 Amazon
Mechanical Turk workers took part. Particip-
ents were rewarded with $10.00.

Design, stimuli, and procedure. The de-
sign, stimuli and procedure were the same as in
Experiment 1, except for the provision of feed-
back. Feedback consisted of the correct answer
being provided to each interpolated testing
question after they submitted their answer. Pro-
vided that the questions were short answer,
feedback consisted of a one word or short sen-
tence providing the correct answer. As the ques-
tions were largely fact-based, no further feed-
back was provided (i.e., why it was the correct
answer). In addition, feedback always occurred
immediately after the participant entered their
response to the interpolated question.

Results

We did not remove any participants in Ex-
periment 2. The raw percentage agreement in
the two markers’ gradings of the short answer
questions was 94% with a Kappa of 0.87. There
was no difference across conditions in terms of
familiarity with the general subject of biochem-
istry, $F(1, 191) = 0.04, p = .84, \eta^2_g = 0.00$, or
how many school courses they had taken in
There was no significant effect of test type, vs. delayed) on final test scores was conducted. The interaction between test type and experiment was not significant, $F(1, 351) = 14.35, p < .001, \eta^2_G = 0.04$, such that participants performed better in the immediate condition and a significant effect of Experiment, $F(1, 351) = 170.1, p < .001, \eta^2_G = 0.47$, $d = 1.89$, such that participants in the immediate condition ($M = 58.3\%, \text{CI } [54.6\%, 62.0\%]$) scored significantly higher than those in the delayed condition ($M = 27.1\%, \text{CI } [24.1\%, 30.1\%]$). We also conducted a two-way between-subjects ANOVA (Experiment: 1 vs. 2) × (Test Type: immediate vs. delayed) on interpolated test scores. There was a significant effect of interpolated test type, $F(1, 351) = 342.75, p < .001, \eta^2_G = 0.49$, such that participants performed better in the immediate interpolated test condition ($M = 83.4\%, \text{CI } [80.2\%, 86.6\%]$) and the delayed interpolated test condition ($M = 81.2\%, \text{CI } [78.1\%, 84.3\%]$) were similar. We also conducted a two-way between subjects ANOVA (Experiment: 1 vs. 2) × (Test Type: immediate vs. delayed) on final test scores. There was a significant effect of interpolated test type, $F(1, 351) = 7.43, p = .007, \eta^2_G = 0.02$, such that participants performed better in the immediate condition and a significant effect of Experiment, $F(1, 351) = 39.97, p < .001, \eta^2_G = 0.10$, such that final test scores were higher in Experiment 2 (with feedback; $M = 82.3\%$) than in Experiment 1 (no feedback; $M = 71.4\%$). The interaction between test type and experiment was not significant, $F(1, 351) = 2.10, p = .15, \eta^2_G = 0.01$.

**Question response time.** As in Experiment 1, we also collected response times to each interpolated and final test question. We report an exploratory analysis here. No participants were removed. In the interpolated tests, an outlier analysis was conducted that excluded response times greater than 2.5 standard deviations from the participant mean, resulting in the removal of 2.2% of the data. A one-way between subject ANOVA (immediate vs. delayed) was conducted on interpolated test response times. There was a significant effect of interpolated test delay, $F(1, 191) = 6.8, p = .01, \eta^2_G = .03, d = .38$, such that participants in the immediate condition ($M = 33,333 \text{ ms, CI } [29,384, 37,282]$) responded faster than participants in the delayed condition ($M = 41,350 \text{ ms, CI } [36,702, 45,999]$). With respect to the final test questions, the outlier analysis resulted in the removal of 3.5% of the data. A one-way between subject ANOVA (immediate vs. delayed) was conducted on final test response times. There was a marginal effect of interpolated test delay, $F(1, 191) = 4.03, p = .05, \eta^2_G = .02, d = .31$, such that participants in immediate condition ($M = 12,187 \text{ ms, CI } [10,095, 14,279]$) responded slower than the participants in the delayed condition ($M = 9,875 \text{ ms, CI } [8,977, 10,773]$).

**Discussion**

Experiment 2 replicated the large effect of test type on performance on the interpolated tests. Participants performed much better when probed immediately after information was presented than approximately three minutes after.
In addition, there was a clear positive effect of feedback on final test scores. Unexpectedly, there was also a negative effect of feedback on interpolated tests scores (we return to this effect in the General Discussion section). Critically, unlike Experiment 1, there was no significant difference between immediate and delayed interpolated test conditions on final test scores. Thus, when feedback is provided, the benefit of immediate over delayed interpolated testing was eliminated. This is consistent with the benefit of immediate interpolated testing (in Experiment 1) reflecting at least in part an increase in the likelihood of successful retrieval and the concomitant re-exposure to the answer. Importantly, the provision of feedback did not lead to a benefit for delayed over immediate interpolated testing. In addition, it is important to note the lack of an interaction between test type and experiment revealed in the combined analysis. This likely reflects a lack of power, provided that the effect in Experiment 1 (\(d = .43\)) decreased from a small to medium sized effect to a small effect in Experiment 2 (\(d = .14\)).

**General Discussion**

Implementing interpolated testing in recorded lectures raises a number of practical questions. In the present experiments, we examined the influence of test lag in the context of interpolated testing in recorded lectures. Specifically, we compared the relative effectiveness of immediate versus delayed interpolated tests both in a context with feedback provided and feedback not provided. Experiment 1 revealed a benefit of immediate and delayed interpolated testing over a no interpolated testing control, thus replicating the beneficial effect of interpolated testing (Szpunar et al., 2013). Unsurprisingly, performance on the interpolated tests was superior when those tests were immediate rather than delayed. On the final test, performance was greater in the immediate interpolated condition than in the delayed interpolated testing condition when no feedback was provided (Experiment 1) but performance did not differ significantly across the two conditions when feedback was provided (Experiment 2). In neither experiment was there any evidence that delayed interpolated testing was superior to immediate interpolated testing. As reviewed in the introduction, recent research relevant to the present work was mixed with respect to whether immediate or delayed interpolated testing would be superior. The present research provides some clarity, at least, within the contexts evaluated. Namely, immediate interpolated testing appears to be superior (or at least equivalent) to delayed interpolated testing and both types of interpolated testing are superior to no interpolated testing.

One critical question that emerges from this work is why delayed testing is not superior to immediate testing (Karpicke & Roediger, 2007). One salient difference between conditions is the large difference in retrieval success across the immediate and delayed interpolated test conditions (in favor of the immediate condition). This can have at least two consequences. First, the higher likelihood of retrieval success in the immediate condition increases the likelihood of re-exposure in that condition (relative to the delayed condition). Re-exposure can be beneficial (e.g., Rawson, Dunlosky, & Thiede, 2000; Roediger & Karpicke, 2006; though not always, see Callender & McDaniel, 2009; Martin, Mills, D’Mello, & Risko, in press; Phillips, Mills, D’Mello, & Risko, 2016). Thus, in Experiment 1, participants were more likely to be re-exposed to the material than individuals in the delayed condition. Experiment 2 supports the idea that differential rates of re-exposure might be operative in the benefit of immediate interpolated testing relative to delayed interpolated testing. In Experiment 2, participants were all provided feedback and, thus, were all re-exposed to the material following each interpolated test. In this sense, feedback could be construed as leveling the playing field in terms of re-exposure across the immediate and delayed interpolated testing conditions. If the higher likelihood of re-exposure to the material in the immediate condition (relative to the delayed condition) during the interpolated tests was partly responsible for the benefit on the final test (in Experiment 1), then equating re-exposure via feedback should have reduced or eliminated the benefit of immediate interpolated testing. This is what we observed in Experiment 2.

While the differential likelihood of re-exposure could explain the lack of a benefit for the delayed interpolated testing condition relative to the immediate interpolated testing condition in Experiment 1, this seems a somewhat unsatisfactory account of why there was no benefit for delayed testing in Experiment 2.
(where re-exposure was more similar). As noted above, in Karpicke and Roediger (2010), provision of feedback led to a benefit for delayed testing over immediate. Such a reversal might have been expected if providing feedback controlled the likelihood of re-exposure across conditions, leaving only the relative benefits of elaborative retrieval (i.e., activation of related information in the course of retrieving the information) which would presumably benefit the delayed condition. There are several potential explanations for the observation of no difference in the feedback condition which are worth considering. The first focuses on a second consequence of the large difference in retrieval success. While feedback reduces the asymmetry in terms of re-exposure to the answer, there still exists a large gap in successful retrieval prior to the provision of that feedback and successful retrieval yields more than just re-exposure. That is, successful retrieval also leads to (or at least has the potential for) elaboration and this elaboration can benefit memory (Carpenter, 2009, 2011). While successful delayed retrieval might involve more elaboration than successful immediate retrieval, the latter likely involves more elaboration than no retrieval (or incorrect retrieval). Thus, the immediate interpolated test condition can be seen as having a greater number of less elaborative retrieval experiences relative to the lower number of more elaborative retrieval experiences in the delayed retrieval condition. From this theoretical perspective, the higher rate of successful retrievals in the immediate interpolated test condition is cancelling out the benefit of greater elaboration in the delayed interpolated test condition. It is also important to note that while the testing that occurred in the immediate condition was much easier than in the delayed condition, it was still difficult (i.e., around 60%–65% correct) relative to contexts that found a benefit of delayed over immediate testing (e.g., Karpicke & Roediger, 2007).

A second explanation emerges from the unexpected difference in interpolated test performance across Experiments 1 and 2. Participants performed better on the interpolated tests overall when they did not receive feedback on those tests. This difference could simply reflect a difference in samples, as Experiments 1 and 2 were conducted at different times. However, this difference might reflect a genuine negative effect of feedback. Indeed, recent work in an anagram task has demonstrated that providing answers reduces persistence (i.e., time spent trying to solve the anagram; Risko, Huh, McLean, & Ferguson, 2017). If participants put less effort into the retrieval of answers to the interpolated tests because the answers are available, whether retrieval was successful or not, then this would likely reduce any potential benefits of elaborative rehearsal and reduce any effect across conditions. The exploratory analysis of question response time does not appear to bear this prediction out. A last potential explanation, and likely the least interesting, is that feedback simply increases performance to such an extent that it reduces the opportunity for genuine differences across conditions to emerge. For example, in Experiment 2 (with feedback), 18% of participants scored 100% on the final test while only 2% of participants did so in Experiment 1 (no feedback). Future work exploring these different explanations for the lack of a difference between immediate and delayed interpolated testing when feedback is provided would be valuable.

Beyond Video Lectures

The focus of the present research has been on the burgeoning use of video lectures and, in particular, an attempt to help inform the use of interpolated testing in that context. That said, interpolated testing is certainly not restricted to recorded lectures. Interpolated questions are also used in live lectures using technologies like clickers and their close cousins (e.g., McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Morling, McAuliffe, Cohen, & DiLorenzo, 2008). Like interpolated testing in a recorded lecture, testing in this context appears to have benefits for retention of lecture material. Live lecture teachers are also presented with a choice as to when to present such quiz questions. Assuming the results here generalize to this live lecture context, the present results could be used to inform decisions for those using “clicker” technology to quiz in class in more traditional brick-and-mortar pedagogical environments.

Limitations

Research in psychology is often criticized for reliance on WEIRD samples (i.e., White, educated, industrialized, rich, democratic; Henrich,
Heine, & Norenzayan, 2010) typically drawn from university courses. The sample used here was drawn from a different and more diverse sample (i.e., Mechanical Turk; Paolacci & Chandler, 2014). While this would usually be considered a strength, if one were concerned only with university samples it might reasonably be considered a limitation. That said, online learning, where video-based lectures are often used, is a format that has the capacity to be accessed by a much broader population than traditional university samples. Indeed, this is one of the most attractive aspects of online education, namely, the asynchronicity affords different populations access to learning opportunities (e.g., Breslow, Pritchard, DeBoer, Stump, Ho, & Seaton, 2013) not typically afforded by traditional brick-and-mortar colleges and universities. Nevertheless, future research examining whether testing effects vary as a function of sample would be valuable, as these methods are increasingly used in live classrooms. In a similar vein, future work focused on different lecture topics would also be of immense value.

Educational Implications

The present research provides important insights with respect to the use of interpolated testing as a pedagogical device. First, we have provided an independent replication of the beneficial effect of interpolated testing in the context of learning complex material. Beyond providing further evidence for the benefit of interpolated testing, the present investigation also provides novel insights into where to place those interpolated tests. Specifically, at least when there is no feedback to be provided, it seems best to place the interpolated tests near to the tested material. When feedback is provided the decision appears less important. This advice seemingly goes counter to the typical practice of only presenting tests at the end of the lecture and appears consistent with recent research on that general topic (Healy et al., 2017; Weinstein et al., 2016; Wissman & Rawson, 2015). It is important to note, however, that much work stands to be done in determining “best practices” for the use of interpolated testing. As noted above, this will include examining potential moderation by population type or learner characteristics in general (e.g., knowledge level) and information type. In addition, there are valuable questions yet to be understood regarding the potential benefit of interpolated testing for long-term retention (e.g., months, years). Lastly, while the content of the lecture used here was not directly on the topic of psychology, there is good reason to expect that the results would generalize to psychology courses. For example, the original work investigating interpolated testing used a statistics lecture (e.g., relation between sample and population; Szpunar et al., 2013, 2014) and here, we used a lecture on biology. Both of which are topics that are not too distantly related to what one would expect in a psychology course (e.g., research methods, physiological psychology). In addition, the testing effect in general has been found across a wide swath of materials (for review see Dunlosky et al., 2013). That said, replicating these results with a traditional psychology lecture would certainly have value in considering the utility of the present work to that particular topic.

Conclusion

In the present investigation, we demonstrated that interpolated testing in a recorded lecture provides significant benefits to information retention over no interpolated testing. We also showed that when no feedback is given, presenting an interpolated test question immediately after information needed to answer the question is given yields further improvements to retention over the alternative of having a delay between this information presentation and interpolated testing. When interpolated test feedback is provided, we found no difference between immediate and delayed interpolated tests. These results have both theoretical and practical implications through furthering our understanding of testing effects and distributed practice as well as providing guidance for educators in their instructional planning.

References


the worldwide classroom: Research into edX’s first MOOC. *Research to Practice in Assessment*, 8, 13–25.


Received September 22, 2017
Revision received July 24, 2018
Accepted July 30, 2018

---

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!