RUNNING HEAD: SPEEDED TESTING

Speeding Lectures to Make Time for Retrieval Practice: Can We Improve the Efficiency of Interpolated Testing?

Evan F. Risko*
Junwen Liu
Laura Bianchi

University of Waterloo

Author Note: This work was supported by funding from the Social Sciences and Humanities Research Council of Canada, the Canada Foundation for Innovation, the Ontario Research Fund, and the Canada Research Chairs program to EFR.

*Corresponding author: 200 University Ave, Waterloo, Ontario, Canada, N2L 3G1, email: efrisko@uwaterloo.ca

CRediT: EFR contributed to Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Visualization, Writing – Original Draft, Writing—review and editing, Supervision, Validation. JL contributed to Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Investigation, Software, Visualization, Writing – Original Draft, Writing—review and editing. LB contributed to Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Investigation, Software, Visualization, Writing—review and editing.

Abstract

Testing is increasingly recognized as an important tool in learning. One form of testing often used in lectures, particularly recorded lectures, is interpolated testing wherein tests are interspersed throughout the lecture. Like testing in general, interpolated testing appears to benefit performance on content tests amongst other outcome variables (e.g., mind wandering). While beneficial, adding testing also increases instructional time. In the present investigation we examine one strategy to mitigate the costs of this increase in instructional time in the context of recorded lectures. Specifically, we examine the interaction between increasing the playback speed of a recorded lecture and adding interpolated tests. Results demonstrate that the conjoint effects of these two interventions is largely additive. That is, the benefit of testing was as robust in a normal speed lecture and a lecture that was sped up 1.5x.

Keywords: testing; learning; playback speed

Public significance statement: Testing has emerged as one of the best means of increasing retention, but it requires added time. In the context of recorded lectures, we examine whether increasing the playback speed of a lecture could reduce the time cost of adding testing without compromising its benefit.

Introduction

Recent research has established testing as an effective tool in improving learning (Agarwal et al., 2021; Dunlosky et al., 2013; Rowland, 2014). Accordingly, integrating testing into existing learning environments (e.g., courses) is generally considered beneficial (Agarwal et al., 2012; 2021). Implementing testing, however, presents its own set of challenges. One of those challenges is that testing adds time to study and instruction. Instructors and students are often faced with teaching or learning a large amount of material. Thus, pedagogical strategies that require added time, even beneficial ones like testing, need to be weighed carefully (Dunlosky & Rawson, 2019). In the present investigation we focus on one type of testing – interpolated testing in recorded lectures – and examine whether a different learning strategy often used to save time – increasing playback speed – can be used to reduce the time costs of the former without eliminating its beneficial influence on learning.

Recorded Lectures

While lecturing has been criticized as a mode of instruction (given its passive nature; Freeman et al., 2014; Theobald et al., 2020), lectures remain a mainstay in education. Recorded lectures, in particular, have long been one of the primary means of communicating educational content in online courses. While recorded lectures might lack some of the magic of attending a live lecture, their benefits are considerable. For example, moving from live to recorded lectures reduces demands on physical space, reduces demands on instructors (e.g., rather than delivering the same lecture repeatedly it can be recorded once and replayed), and offers students the ability to individually control lecture delivery (i.e., they can start, stop, replay, and view the material at their convenience). In addition, recorded lectures are arguably more accessible. For example, advances in technology now allow automatic closed captioning of recorded lectures that in some cases permit translation of the recorded lecture into other languages. Lastly, recorded lectures, once largely relegated to online courses, have been given new life as a critical tool in the context of flipped classrooms where students are provided with and expected to consume the recorded lecture content outside of class thus allowing in-class time to be devoted to active learning activities (e.g., Akçayır & Akçayır, 2018). Thus, the recorded lecture might someday replace the live lecture as the most used format for the communication of lecture-based content. In a similar vein, with the COVID-19 pandemic requiring a large-scale shift to online education and as a result the exposure of stakeholders (i.e., institutional administrators, instructors, students) to the many benefits of recorded lectures, their use will likely increase dramatically. As a result, research aimed at developing effective recorded lectures stands to have a large impact on education moving forward.

Interpolated Testing

One means of improving learning from recorded lectures is to interpolate test questions into them (Haagsman et al., 2020; Jing et al., 2016; Lavigne & Risko, 2018; Szpunar et al., 2013, 2014; van der Meij & Böckmann, 2021; Yang et al., 2020). For example, the recorded lecture might stop, a question is posed to the learner, the learner answers the question and possibly receives feedback, then the lecture continues. While research on interpolated testing is nascent, there exist a number of demonstrations that it might represent a useful strategy in the design of recorded lectures. In one of the first examinations of interpolated testing in lectures, Szpunar et al. (2013) presented participants with a statistics lecture broken into 4 segments. Across two experiments, each of the first 3 segments were followed by a math task after which participants either received a test based on the previous segment, continued to perform the math task, or restudied the material (i.e., were provided questions and answers). Results demonstrated that those in the interpolated testing condition outperformed the other two conditions both overall and on the segment 4 questions. Critically, the latter items (unlike the items from segments 1-3) had not been previously tested (i.e., this was the first time the participants had seen those questions). Thus, interpolated testing appears able to produce both backward (i.e., testing enhances long term retention) and forward (i.e., testing enhances new learning) testing effects (e.g., Szpunar et al., 2008; Yang et al., 2018, 2021). Subsequent research on interpolated testing has typically found at least some form of learning benefit but not always (Haagsman et al., 2020; Jing et al., 2016; Lavigne & Risko, 2018; Szpunar et al., 2014; van der Meij & Böckmann, 2021; Welhaf et al., 2022).

As noted above, interpolated testing appears to produce both backward and forward testing benefits in the context of lectures. There exist a number of mechanisms that have been postulated to account for these effects. For example, with respect to the backward testing effect, an influential account – the episodic context account (Karpicke et al., 2014; Lehman et al., 2014) – attributes the benefit to testing leading to the storage of additional contextual cues, thus increasing the likelihood of future recall. A number of other accounts are also available for the backward testing effect (e.g., Carpenter, 2011; Rowlands, 2014). With respect to the forward testing effect, Yang et al. (2021) discuss four mechanisms (1) testing reduces proactive interference by inducing context changes, (2) testing leads to a more effective strategy, (3) testing resets the encoding process leaving greater capacity for subsequent encoding, and (4) testing leads individuals to invest more effort (see Yang et al., 2018 for other potential mechanisms). Yang et al. (2021) provides compelling evidence for (1; but see Ahn & Chan, 2022) and (2) but not (3) and did not test (4). Evidence for the latter (i.e., testing leads to increases in effort) was provided by Yang et al. (2019) who demonstrated that study time decreased with each list in a setting with no tests but stayed more constant in a setting with tests. Thus, if we interpret the amount of study time an individual voluntarily allocates to be an index of effort, individuals invested more effort in the condition with tests.

Interpolated testing also appears to influence a number of other learning relevant variables though this research has been more limited to date and less consistent. For example, in the original Szpunar et al. (2013) investigation they demonstrated that individuals in the interpolated test condition reported mind wandering less and took more notes than those in the other conditions. However, those effects have been less consistent (Jing et al., 2016; Szpunar et al., 2014; Welhaf et al., 2022). Szpunar et al. (2013) also found that individuals in their interpolated test condition reported that the experience of learning the lecture material was less “mentally taxing,” led to less anxiety about an upcoming test, and produced less negative affect than individuals in their restudy and non-tested group. They found no effect on positive affect. In a subsequent study, Szpunar et al. (2014), using a design similar to Szpunar et al. (2013), found that interpolated testing boosted metacognitive calibration (i.e., the absolute difference between performance estimates and actual performance was smaller). While the majority of work to date has not allowed participants to interact with the video lecture (i.e., rewind, stop, replay), recent work that permitted this control has found that interpolated testing also influences how individuals interact with recorded lectures. Haagsman et al. (2020) found that lectures with interpolated tests (pop-up questions) were rewound more often in the 30 seconds prior to the questions but rewound and fast-forwarded less overall than those without interpolated tests. In a similar vein, van der Meij and Böckmann (2021) found that lectures with interpolated tests were replayed more often and for longer, which might suggest more engagement. Taken together, research investigating interpolated testing in recorded lectures suggests that the net effect on learning is beneficial. In addition, interpolated testing appears to influence a number of other learning relevant variables including attention, mental demand, affect, metacognition, and style of interaction. That said, much remains to be understood about its influence and the best means of implementing interpolated testing in practice.

All else being equal, adding interpolated tests to a recorded lecture requires increasing their overall duration. This increase will be equal to the amount of time required to read and answer the interpolated questions (and feedback if provided) and as such will vary with the number and type of questions asked, amongst other variables (e.g., how much effort learners invest in answering the questions). At the most basic level, increasing the duration of the lecture reduces the time available to devote to other learning activities. This opportunity cost when considered at the level of an individual lecture might be considered small, but across an entire course (and across multiple courses) this time cost would be nontrivial. As noted above, individual learner's time is often limited, and implementing practices that take additional time should be done with great care. In addition to increasing time demands, increasing the duration of a lecture might have other negative effects. One salient potential cost is that increasing the duration of the lecture increases the amount of time individuals need to sustain attention, and sustaining attention for long periods of time can be challenging in general (Grier et al., 2003; Manly et al., 1999). This appears to apply to recorded lectures as well (Farley et al., 2013; Guo et al., 2014; Risko et al., 2012; Seli et al., 2016).

One potential strategy to address time costs introduced by adding interpolated tests to recorded lectures is to increase the playback speed of the recorded lecture. Individuals appear able to comprehend speech at rates much higher than natural speech (see Barron, 2013, for a review). In addition, recent research has demonstrated that this is also true in the context of recorded lectures (see Pastore & Ritzhaupt, 2015). For example, Wilson and colleagues (2018) demonstrated that increasing the playback speed of a lecture 1.6-1.7x lead to either no (Experiment 1) or a small (Experiment 2) performance cost on a test of the lecture material. Similarly, Nagahama and Morita (2018; see also Nagahama & Morita, 2017) increased the playback speed of a lecture 1.5x and 2x and found no significant performance cost on a test of the lecture material relative to the original speed. That said, increasing the playback speed of lectures does have costs when the playback speeds increase beyond more modest levels (Jacobson et al., 2018; Murphy et al., 2022). For example, Murphy et al., (2022) compared lecture speeds of 1x, 1.5x, 2x, and 2.5x and found no performance costs up to 2x but a clear performance cost at 2.5x. In addition, how increasing playback speed might interact with other variables is largely unexplored. Nevertheless, a modest increase in playback speed could produce enough savings in terms of instructional time to counteract the increase in lecture duration caused by the introduction of interpolated tests, thus improving the overall efficiency of the interpolated testing approach. We examine this idea here.

Present Investigation

In the present investigation individuals viewed a lecture composed of 4 segments that featured interpolated tests or no tests and was presented at an increased speed or the recorded speed. Specifically, two lectures included interpolated tests following each of the first three segments and no interpolated test following the fourth segment. Critically, one of the two lectures with interpolated tests was presented at the original speed and the other was sped up approximately 1.5x the original speed. This increase in playback speed reduced the total lecture length to approximately match the length of the lecture with no interpolated tests. Thus, we can compare two lectures that are the approximately the same duration, contain the same content, but one features interpolated tests and the other does not. Participants completed a test based on the lecture material following the lecture that included questions from all four segments. The questions from segments 1-3 were the same as the interpolated tests and the questions from segment 4 had not been answered previously. As described above, the contrast between these two question types is important theoretically as the latter are not tested, hence performance on these items provide a test of forward testing effects. In addition to assessing the full factorial design crossing playback speed and interpolated testing, we provided a direct comparison of the normal lecture with no tests and the speeded lecture with interpolated tests. This contrast may be useful practically in the sense that it features a comparison of lectures of equivalent length (i.e., one without interpolated tests, and one with tests sped to account for the added time).

In addition to examining the influence of interpolated testing on memory for lecture material, we also sought to further examine its influence on participant's metacognitions. To this end, participants provided an estimate of their test performance both prior to the test (prospective) and following the test (retrospective). As noted above, Szpunar et al. (2014) found that interpolated testing improved participant's prospective calibration. The present investigation tests this again, examines its potential moderation by lecture speeding, and extends it by adding retrospective judgements (i.e., how well they believed they did after completing the test). The latter allow us to test one explanation of the benefit of interpolated testing on prospective calibration. Specifically, one explanation of this benefit is that it is based on the experience of test taking. That is, individuals in the interpolated testing condition have experience taking the kind of test for which they are being asked to estimate their performance whereas individuals in the other condition are not. This experience could inform their judgements and improve calibration. If this is the case, then we would expect that the effect would be reduced or eliminated in the case of retrospective judgements because the latter come after the final test, hence all groups will have been exposed to a test prior to making this judgement.

We also examined the influence of interpolated testing and increasing playback speed on subjective effort. In theoretical accounts of multimedia learning, cognitive load is often held to represent a central construct (e.g., Mayer & Moreno, 1998, 2003; Paas et al., 2003; Sweller, 2011). Learners are considered to have a limited capacity, and instructional design that optimizes the use of that limited resource are held to optimize outcomes. Both interpolated testing and increasing playback speed should increase load. With respect to subjective effort, we used items indexing both required effort and invested effort (Koriat & Nussinson, 2009). These items can map roughly onto the distinction between difficulty and effort (Paas & Van Merrienboer, 1994). While invested and required effort are typically related, they can be dissociated (Koriat & Nussinson, 2009; Van Kessel, Ashburner & Risko, in press). Both interpolated testing and increasing lecture speed could increase load. The inclusion of effort measures provides a direct test of this notion in terms of subjective effort (Paas & Merrienboer, 1994) and the potential interaction between the two manipulations.

Another important consideration in using increased playback speed to make time for interpolated testing is how this strategy impacts affect and/or the lecture experience in general (Plass & Kalyuga, 2019). For example, in the Cognitive Affective Theory of Multimedia Learning (CATML; Moreno & Mayer, 2007), changes in affect can influence motivation which impacts learning through its effect on cognitive engagement. While this link between motivation and effort investment in instructional contexts has long been recognized (Paas et al., 2005), examining how the design of recorded lectures (e.g., speeding, interpolated testing) impacts such variables has arguably lagged behind. These considerations are particularly important in the present context given that increasing playback speed has, in some cases, been associated with increases in negative affect or decreases in “liking” aspects of the learning experience (Wilson et al., 2018). This negative experience could undermine learner’s motivation to invest effort. In a similar vein, a negative experience could produce more direct effects by increasing cognitive load. For example, increasing playback speed might increase performance anxiety, drawing resources away from the lecture toward emotion regulation (e.g., Plass & Kalyuga, 2019). Thus, we also measured a number of variables related to affect. We measured participants' liking of the lecture, their likelihood of watching a similar lecture in the future, and their experience of positive and negative affect using the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). The latter can be considered a measure of "core" affect while the former measures provide an index of "attributed" affect (Russell, 2003).

Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. Data, pre-registrations, and analysis code for the reported analyzes are available at [E1 pre-registration: https://osf.io/m8zc9; Data and analysis code: https://osf.io/hmjta; Liu et al., 2023; E2 pre-registration: https://osf.io/c785z; Data and analysis code: https://osf.io/254pt; Lab, 2023]. Research materials are available upon request. Both Experiments 1 and 2 received ethics clearance.

Experiment 1[1]

Method

Participants

A total of 362 participants completed the study. There was a total of 321 participants who satisfied the inclusion criteria (see results section for details about exclusions). The target sample size was 80 participants per condition determined by rounding up a power analysis with F = .165, α = .05, power = .80, df = 1, and four groups using G*Power, which yielded 73 participants per cell. Participants were recruited from the Prolific online recruiting platform, and they were given 5.25 British Pounds as compensation for their participation. Mean age was 36 years, 51% selected female, 47% male, and 2% other, and the mean number of online courses taken was 7.6 with the median 4. The number of online courses required text entry. If participants put in a value that was not a number (e.g., “many”, or “20+), then we converted to a number if it was reasonable (e.g., convert 20+ to 20) or removed the data point if not (e.g., “many”). As the difference in the mean and median suggests, the distribution was positively skewed with several participants having taken many online courses but most taking a few. In terms of education, most reported having a bachelor’s degree (41%) with 40% having less and 18% more than a bachelor’s degree. For reading proficiency 94% selected high proficiency, and for listening proficiency 92% selected high proficiency.

Design

We used a 2 (interpolated tests: test vs. no tests) x 2 (speed: normal vs. speeded) between participants design. Table 1 includes condition descriptions.

Table 1. Descriptions of each of the conditions.

Speed	Interpolated Testing	Description
Normal	No Interpolated Tests	Participants viewed a lecture presented at the originally recorded speed without interpolated tests
1.5x	No Interpolated Tests	Participants viewed a lecture with the playback speed increased 1.5x without interpolated tests
Normal	Interpolated Tests	Participants viewed a lecture presented at the originally recorded speed with interpolated tests
1.5x	Interpolated Tests	Participants viewed a lecture with the playback speed increased 1.5x with interpolated tests

Materials

Demographics. Participants were asked to provide their age, gender, education level, and number of online courses they had taken. Participants were also asked to indicate their level of proficiency in reading and listening in English on 4-point scales, with 1 being no proficiency, and 4 being high proficiency.

Lecture. The video lecture used was the same as that used in Szpunar et al. (2013)[2]. The video lecture lasted 21 minutes and 35 seconds and consisted of a typical university lecture, covering introductory concepts in statistics (e.g., samples, population, descriptive statistics, random sampling, study types, confounds, causation). Pausing and rewinding were not allowed. The video is a recording of the lecture slides, taken from a front view angle, accompanied by the voice of a male lecturer. The video was split into four segments of roughly equal length, each being just under 5.5 minutes long. The speeding manipulation reduced the length of each segment to under 3.7 minutes, saving roughly 110 seconds.

Test questions. The same 24 short answer test questions, six for each segment, used in Szpunar et al. (2013) were used here and were retrieved from the first author of that study. Questions addressed material in the lecture and typically required a single word response (e.g., “Define N and n,” “Confounding variables can be controlled in this type of study”). In the interpolated testing conditions, the 18 questions for segments 1, 2, and 3 were shown immediately after each respective segment. Participants had 110 seconds to complete each set of six questions. Questions for segment 4 only appeared in the cumulative test. The cumulative test contained all 24 questions presented in a randomized order. The cumulative test was self-paced. Each answer was scored by two naive coders as either correct or incorrect. The correlation between coders on the cumulative test was $r = .96$. Disagreements were adjudicated by author JL.

Metacognition. Before completing the cumulative test, participants were asked to estimate, using a percentage scale, how well they would do on the test. After the test, participants were asked to estimate how well they had done on the test. We used the same measure of calibration as that reported by Szpunar et al. (2014) in their demonstration that interpolated testing benefits prospective calibration. This measure is derived from participant's performance estimate and their actual performance. The measure puts the absolute difference between participant's predicted and actual performance on a scale between 0 (poor) and 100 (perfect).

$$\left(1 - \frac{|Estimate - Actual|}{100}\right) x\ 100$$

Effort. Effort was measured by two 9-point scales designed to index data-driven effort and goal-driven effort. The data-driven effort question asked how much effort the participants thought the lecture required. The goal-driven effort question asked how much effort participants chose to invest. These scales ranged from 1 (*very, very little effort*) to 9 (*very, very much effort*).

Affect. Participants' impression of the lecture was measured using two 7-point scales. The first scale asked participants how much they "liked" the lecture, ranging from 1 (*very much disliked*) to 7 (*very much liked*). The second scale asked participants how likely they would be to watch a video lecture in the same format in the future, ranging from 1 (*very unlikely*) to 7 (*very likely*). Positive and negative affect were measured using the PANAS (Watson et al., 1988). The inventory consists of twenty 5-point scales asking participants the degree to which they experienced various mood terms during the lecture (1 = *not at all* to 5 = *a lot*).

Attention Checks. To ensure participants were paying adequate attention, two attention check questions were included. One question was presented among the cumulative test questions, and it asked whether 9 or 34 was numerically larger. The other question appeared among the PANAS questions, simply requiring participants to pick option 4 on the 5-point scale. Only data from participants who correctly answered both attention checks were included in the analyses.

Post-Study Survey. At the end of the experiment participants were asked whether they Googled answers to the test, took notes, and watched the lecture fully. In addition, they were asked whether there was a reason we should not use their data. The latter question was included to provide participants an opportunity (without risk of losing payment) to communicate reasons that we should not use their data. Participants admitting to Googling answers, taking notes, not watching the lecture fully or providing a reason that we should not use their data were excluded.

Procedure

The core procedure mirrored that of Szpunar et al. (2013). Once the participant signed up on Prolific, they were provided a link to the study, presented on Qualtrics. After passing a bot check and providing consent, the participant answered the demographic questions and read the instructions for the study phase.

The study phase involved participants watching the video lecture broken into four segments with a 30-second break after the first, second and third segments. In the normal speed conditions participants viewed the lecture at its recorded speed, and in the increased speed condition participants viewed the lecture segments at 1.5x the recorded speed. In the interpolated testing condition participants answered the six questions after each segment. They had 110 seconds to complete these questions. Participants could not proceed until the 110 seconds had elapsed. This was followed by a 30-second break. Participants were not provided with any instructions about what to do during the break. In the no interpolated testing condition there were no tests after each segment. After the study phase, participants provided their performance estimate, then began the cumulative test. The test was self- paced and after completing the test participants estimated how well they did, then answered the affect and effort-related questions. Lastly, participants were asked whether they searched for answers online, took notes, or failed to watch the lecture fully. Participants were also asked to provide any reasons why their data should not be used. Depending on the condition, the entire study session took between 30 to 50 minutes.

Results

As noted above, a total of 362 participants completed the study and 321 participants satisfied the inclusion criteria. Data from 5 participants were excluded for providing a reason not to use their data, 3 failed an attention check, 10 Googled the answers, 11 took notes, 4 participants mentioned not watching the lecture fully, and 8 were excluded for more than one reason. There were 82 participants in the normal speed with no interpolated tests condition, 80 in the increased speed with interpolated tests condition, 80 in the normal speed with interpolated tests condition, and 79 in the increased speed with no interpolated tests condition. A 2 (interpolated tests: test vs. no tests) x 2 (speed: normal vs. speeded) between- subjects ANOVA was conducted for each dependent variable. In addition, we directly compared the normal speed with no interpolated tests and the increased speed with interpolated tests conditions. These conditions took a similar amount of time, but the latter included interpolated tests and increased playback speed in order to compensate for their added time. ANOVAs using a White adjustment, Welch's t-tests, and Wilcox Rank Sum Tests (assessing the main effects of testing and increasing playback speed) were also conducted to address assumption violations (e.g., negative PANAS was heavily positively skewed, and calibration negatively skewed) and were qualitatively similar to the reported results. Means and bootstrapped confidence intervals for each dependent variable, for each condition, are presented in Table 2. The correlations between the dependent variables in each condition are provided in Supplementary Materials Table S1-S4.

Table 2. Means and 95% confidence intervals (bootstrapped) for each dependent variable and condition in Experiment 1.

Measurements	Standard Speed - No Test			Standard Speed - Test			Increased Speed - No Test			Increased Speed - Test		
	M	95% CI		*M*	95% CI		*M*	95% CI		*M*	95% CI	
		LL	*UL*		*LL*	*UL*		*LL*	*UL*		*LL*	*UL*
Test Accuracy												
Segment 1, 2, & 3	**46.1**	*41.5*	*50.4*	**56.9**	*50.9*	*62.6*	**46.8**	*42.7*	*51.1*	**56.9**	*52.0*	*61.6*
Segment 4	**37.8**	*32.1*	*44.1*	**46.0**	*39.2*	*52.9*	**35.4**	*30.6*	*40.3*	**51.5**	*45.4*	*57.5*
Interpolated Tests				**58.5**	*53.1*	*63.9*				**58.0**	*53.6*	*62.3*
Metacognition												
Prospective Estimate	**62.0**	*57.1*	*66.1*	**57.9**	*52.4*	*63.1*	**63.5**	*58.8*	*67.6*	**60.3**	*54.3*	*65.6*
Prospective Calibration	**78.3**	*74.9*	*81.3*	**82.2**	*78.6*	*85.1*	**74.4**	*70.4*	*77.8*	**82.4**	*79.0*	*85.3*
Retrospective Estimate	**36.1**	*31.3*	*41.1*	**47.8**	*41.1*	*54.0*	**38.3**	*33.3*	*43.6*	**51.5**	*45.5*	*56.7*
Retrospective Calibration	**83.8**	*81.5*	*85.9*	**83.9**	*80.3*	*86.8*	**83.8**	*80.0*	*87.2*	**85.1**	*82.2*	*87.7*
Effort												
Goal- Driven	**6.72**	*6.35*	*7.06*	**7.19**	*6.85*	*7.48*	**7.39**	*7.08*	*7.66*	**7.59**	*7.25*	*7.86*
Data- Driven	**6.56**	*6.18*	*6.91*	**7.07**	*6.72*	*7.37*	**6.98**	*6.58*	*7.30*	**7.22**	*6.83*	*7.54*
Affect												
Liking	**4.72**	*4.35*	*5.06*	**4.06**	*3.68*	*4.41*	**4.79**	*4.43*	*5.11*	**4.46**	*4.05*	*4.84*
Future Lecture	**4.49**	*4.04*	*4.85*	**3.52**	*3.06*	*3.96*	**4.58**	*4.15*	*4.93*	**3.89**	*3.48*	*4.28*
Positive PANAS	**2.50**	*2.34*	*2.69*	**2.52**	*2.35*	*2.73*	**2.66**	*2.49*	*2.81*	**2.82**	*2.64*	*2.99*
Negative PANAS	**1.35**	*1.26*	*1.48*	**1.43**	*1.32*	*1.58*	**1.32**	*1.23*	*1.44*	**1.58**	*1.44*	*1.76*

Test Performance

There was a main effect of testing on the combined accuracy of segments 1, 2, and 3, $F(1, 317) = 17.72$, $p < .001$, $\eta^2_G = .05$, such that testing led to higher accuracy. There was no main effect of increasing playback speed, $F(1, 317) < 1$, $p = .87$, $\eta^2_G < .01$, nor an interaction, $F(1, 317) < 1$, $p = .89$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 3.24$, $p = .001$, $d = 0.51$, such that performance was greater in the latter condition.

There was a main effect of testing on segment 4 accuracy, $F(1, 317) = 15.44$, $p < .001$, $\eta^2_G = .05$, such that testing led to significantly higher accuracy. There was neither a main effect of increasing playback speed, $F(1, 317) < 1$, $p = .62$, $\eta^2_G < .01$, nor an interaction, $F(1, 317) = 1.59$, $p = .21$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 3.10$, $p = .002$, $d = 0.49$, such that performance was greater in the latter condition.

On the interpolated tests themselves there was no effect of speeding, $t(158) = 0.16$, $p = .87$, $d = 0.02$.

Metacognition

There was no main effect of testing, $F(1, 317) = 1.94$, $p = .16$, $\eta^2_G = .01$, increasing playback speed, $F(1, 317) < 1$, $p = .45$, $\eta^2_G < .01$, or interaction, $F(1, 317) < 1$, $p = .86$, $\eta^2_G < .01$, on prospective performance estimates. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(160) = 0.45$, $p = .66$, $d = 0.07$.

There was a main effect of testing on retrospective performance estimates, $F(1, 317) = 19.09$, $p < .001$, $\eta^2_G = .06$, such that performance estimates were higher with interpolated testing. There was no main effect of increasing playback speed, $F(1, 317) < 1$, $p = .30$, $\eta^2_G < .01$, or interaction, $F(1, 317) < 1$, $p = .78$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 4.05$, $p < .001$, $d = 0.64$, such that performance estimates were greater in the latter condition.

With respect to calibration, there was a main effect of testing, $F(1, 317) = 12.40$, $p < .001$, $\eta^2_G = .04$, on prospective calibration, such that participants were better calibrated in lectures with interpolated tests. There was no main effect of increasing playback speed, $F(1, 317) = 1.18$, $p = .28$, $\eta^2_G < .01$, or interaction, $F(1, 317) = 1.45$, $p = .23$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(160) = 1.80$, $p = .07$, $d = 0.28$.

There was no main effect of testing, $F(1, 317) < 1$, $p = .67$, $\eta^2_G < .01$, increasing playback speed, $F(1, 317) < 1$, $p = .64$, $\eta^2_G < .01$, or interaction, $F(1, 317) < 1$, $p = .70$, $\eta^2_G < .01$, on retrospective calibration. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(160) = 0.73$, $p = .46$, $d = 0.12$.

Effort

There were main effects of both testing, $F(1, 317) = 10.86$, $p = .001$, $\eta^2_G = .03$, and increasing playback speed, $F(1, 317) = 4.30$, $p = .04$, $\eta^2_G = .01$, on goal-driven effort, such that participants invested more effort when the lecture had interpolated tests and when it was speeded. There was no interaction, $F(1, 317) < 1$, $p = .40$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 3.63$, $p < .001$, $d = 0.57$, such that reports were higher in the latter condition.

There was no main effect of testing on data-driven effort, $F(1, 317) = 2.48$, $p = .12$, $\eta^2_G = .01$. There was a main effect of increasing playback speed, $F(1, 317) = 4.23$, $p = .04$, $\eta^2_G = .01$, such that participants found the lecture to require more effort when it was speeded. There was no interaction, $F(1, 317) < 1$, $p = .49$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 2.56$, $p = .01$, $d = 0.40$, such that reports were higher in the latter condition.

Affect

There was no main effect of testing on liking, $F(1, 317) = 1.58$, $p = .21$, $\eta^2_G < .01$. There was a main effect of increasing playback speed on liking, $F(1, 317) = 6.95$, $p = .01$, $\eta^2_G = .02$, such that participants who watched the lecture with increased playback speed had a lower rating on the liking measure. There was no interaction, $F(1, 317) < 1$, $p = .37$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(160) = 0.96$, $p = .34$, $d = 0.15$.

There was no main effect of testing on the likelihood of watching another lecture in the same format, $F(1, 317) = 1.14$, $p = .29$, $\eta^2_G < .01$. There was a main effect of increasing playback speed, $F(1, 317) = 15.12$, $p < .001$, $\eta^2_G = .05$, such that participants who watched the speeded lecture reported being less likely to watch another lecture in the future in the same format. There was no significant interaction, $F(1, 317) < 1$, $p = .44$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 2.00$, $p = .046$, $d = 0.32$, such that the likelihood was higher in the former condition.

There was a main effect of testing on positive PANAS, $F(1, 317) = 6.28$, $p = .01$, $\eta^2_G = .02$, such that scores were higher in the lecture with interpolated tests. There was no main effect of increasing playback speed, $F(1, 317) < 1.01$, $p = .31$, $\eta^2_G < .01$, or interaction, $F(1, 317) < 1$, $p = .43$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 2.49$, $p = .01$, $d = 0.39$, such that positive affect scores were higher in the latter condition.

There was no main effect of testing on negative PANAS, $F(1, 317) < 1$, $p = .34$, $\eta^2_G < .01$. There was a main effect of increasing playback speed, $F(1, 317) = 6.83$, $p = .009$, $\eta^2_G = .02$, such that negative affect scores were higher for the speeded lecture. There was no interaction, $F(1, 317) = 1.94$, $p = .16$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(160) = 2.34$, $p = .02$, $d = 0.37$, such that negative affect scores were higher in the latter condition.

Discussion

The Experiment 1 results were straightforward. We replicated the beneficial effect of interpolated testing on items that had been tested (segment 1, 2, and 3) and those that had not (segment 4). Increasing playback speed did not impair performance on memory for the lecture material. When we consider the lectures that took an equivalent amount of time, the lecture that was speeded with interpolated tests yielded better performance than the normal speed lecture that had no interpolated tests. Thus, if we consider only test performance, then increasing playback speed seems a potential tool that one could employ to compensate for the added time needed for interpolated testing.

Moving beyond memory for lecture material, across all of the measures here we found no interaction between the inclusion of interpolated tests and increasing playback speed. However, we did find a number of main effects that provide additional insight into these learning strategies. Starting with interpolated testing, participants' prospective performance estimates were unaffected by the presence of tests, but their prospective calibration increased. These results were generally consistent with Szpunar et al. (2014). After individuals experienced the test, this pattern changed. Individuals' retrospective estimates were higher in the condition with interpolated tests and there was no difference in calibration. Thus, experience with the test brought performance estimates in greater alignment with actual performance. This is consistent with an explanation of the prospective benefit in calibration in the interpolated testing condition being a function of experience test taking. With respect to affect, the inclusion of interpolated tests had limited effects. There was an increase in positive affect, as measured by the PANAS, when interpolated tests were included. Lastly, with respect to effort, individuals reported investing more effort in the lectures with interpolated tests but reported they did not require more effort.

Turning to increasing playback speed, there was no effect on the metacognitive variables. Increasing playback did increase subjective effort both in terms of what learners reported investing and what learners perceived to be required. In addition, increasing playback speed impacted measures of affect. Individuals reported not liking the increased speed lectures as much, not being as willing to watch another lecture in the increased speed format and experiencing more negative affect while viewing that lecture. Thus, increasing playback speed appears to have led to a less positive learning experience. Before considering the implications of these results we report the results of a second experiment replicating and extending the results of Experiment 1.

Experiment 2

In Experiment 2 we sought a replication of Experiment 1 with a slight modification. Specifically, in Experiment 1 individuals responded to items related to affect and effort after having completed the cumulative test and associated metacognitive questions. This raises the possibility that performance on the test influenced the affect and effort reports. This could be considered particularly concerning provided the sizeable difference in performance across the testing groups. In Experiment 2 we replicate Experiment 1 and shift the affect and effort items to be prior to the final test and associated metacognitive questions.

Method

Participants

A total of 355 participants completed the study. There was a total of 312 participants who satisfied the inclusion criteria (see results section for details about exclusions). Inclusion criteria, determination of sample size, and compensation were the same as in Experiment 1. Mean age was 39 years, 61% selected female, 38% male, and 1% other, and the mean number of online courses taken was 5.1 with the median 2. In terms of education, most reported having a bachelor's degree (43%) with 38% having less and 19% more than a bachelor's degree. In reading proficiency 95% selected high proficiency, and for listening proficiency 93% selected high proficiency.

Materials and Procedure

The Design, Materials and Procedure were identical to Experiment 1 except that the affect, and effort-related questions were moved to before the cumulative test and associated metacognitive prompts. In addition, provided the high correlation between naïve coders in Experiment 1, only a single naïve coder was used in Experiment 2.

Results

As noted above, 355 participants were collected, and 312 met the inclusion criteria. Data from 9 participants were excluded for providing a reason not to use their data, 9 failed an attention check, 7 Googled the answers, 9 took notes, 3 participants mentioned not watching the lecture fully, and 6 were excluded for more than one reason. There were 78 participants in the normal speed with no interpolated tests and increased speed with no interpolated tests conditions, 76 in the normal speed with interpolated tests condition, and 80 in the increased speed with interpolated tests condition. A 2 (interpolated tests: test vs. no tests) x 2 (speed: normal vs. speeded) between-subjects ANOVA was conducted for each dependent variable. The statistical approach was the same as in Experiment 1. Means and bootstrapped confidence intervals for each dependent variable, for each condition, are presented in Table 3. The correlations between the dependent variables in each condition are provided in Supplementary Materials Table S5-S8.

Table 3. Means and 95% confidence intervals (bootstrapped) for each dependent variable and condition in Experiment 2.

Measurements	Standard Speed - No Test			Standard Speed - Test			Increased Speed - No Test			Increased Speed - Test		
	M	95% CI		*M*	95% CI		*M*	95% CI		*M*	95% CI	
		LL	*UL*		*LL*	*UL*		*LL*	*UL*		*LL*	*UL*
Test Accuracy												
Segment 1, 2, & 3	**47.2**	*42.5*	*51.8*	**57.7**	*52.4*	*62.2*	**37.3**	*32.4*	*42.3*	**53.4**	*48.4*	*58.6*
Segment 4	**36.1**	*31.1*	*41.2*	**48.9**	*42.3*	*55.5*	**31.8**	*26.5*	*37.0*	**44.0**	*38.3*	*50.0*
Interpolated Tests				**60.0**	*54.8*	*64.8*				**54.0**	*49.2*	*58.6*
Metacognition												
Prospective Estimate	**58.9**	*54.1*	*64.6*	**53.1**	*48.2*	*57.4*	**54.4**	*49.3*	*59.5*	**50.2**	*45.2*	*55.3*
Prospective Calibration	**77.6**	*73.7*	*81.3*	**83.4**	*80.3*	*86.0*	**77.3**	*73.5*	*80.7*	**84.6**	*81.2*	*87.3*
Retrospective Estimate	**38.4**	*33.2*	*43.6*	**44.7**	*39.5*	*49.7*	**30.8**	*25.8*	*36.3*	**41.4**	*36.0*	*46.8*
Retrospective Calibration	**84.1**	*81.2*	*86.6*	**83.0**	*80.2*	*85.5*	**86.1**	*83.8*	*88.2*	**84.8**	*82.2*	*87.1*
Effort												
Goal-Driven	**7.11**	*6.73*	*7.42*	**7.30**	*6.92*	*7.61*	**7.55**	*7.19*	*7.82*	**8.00**	*7.73*	*8.20*
Data- Driven	**6.14**	*5.73*	*6.52*	**6.71**	*6.30*	*7.09*	**6.94**	*6.47*	*7.33*	**7.54**	*7.15*	*7.81*
Affect												
Liking	**4.78**	*4.46*	*5.08*	**4.50**	*4.11*	*4.84*	**4.24**	*3.81*	*4.65*	**4.08**	*3.66*	*4.44*
Future Lecture	**4.50**	*4.09*	*4.88*	**4.28**	*3.79*	*4.72*	**3.99**	*3.52*	*4.42*	**3.78**	*3.34*	*4.20*
Positive PANAS	**2.74**	*2.58*	*2.92*	**2.70**	*2.51*	*2.90*	**2.77**	*2.60*	*2.96*	**2.85**	*2.66*	*3.04*
Negative PANAS	**1.14**	*1.09*	*1.24*	**1.19**	*1.13*	*1.29*	**1.46**	*1.34*	*1.62*	**1.48**	*1.37*	*1.61*

Test Performance

There was a significant main effect of testing on the combined accuracy of segments 1, 2, and 3, $F(1, 308) = 28.23$, $p < .001$, $\eta^2_G = .08$, such that testing led to higher accuracy. There was also a significant main effect of increasing playback speed, $F(1, 308) = 8.07$, $p = .005$, $\eta^2_G = .03$, such that those in the speeding condition performed more poorly. There was no interaction, F(1,

308) = 1.30, p = .26, η^2_G < .01. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(156) = 1.76$, p = .08, d = .28.

There was a significant main effect of testing on segment 4 accuracy, $F(1, 308) = 17.78$, p < .001, η^2_G = .05, such that testing led to higher accuracy. There was neither a significant main effect of increasing playback speed, $F(1, 308) = 2.44$, p = .12, η^2_G = .01, nor a significant interaction, $F(1, 308) < 1$, p = .91, η^2_G < .01. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(156) = 1.91$, p = .06, d = 0.30.

On the interpolated tests themselves there was no effect of speeding, $t(154) = 1.71$, p = .09, d = 0.27.

Metacognition

There was a significant main effect of testing on prospective performance estimates, $F(1, 304) = 4.55$, p = .03, η^2_G = .01, such that participant's estimates were higher in the no interpolated tests condition. There was no significant main effect of increasing playback speed, $F(1, 308)$ = 2.70, p = .10, η^2_G = .01, or an interaction, $F(1, 316) < 1$, p = .71, η^2_G < .01, on prospective performance estimates. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 2.65$, p = .008, d = 0.42, such that prospective performance estimates were higher in the former condition.

There was a significant main effect of testing on retrospective performance estimates, $F(1, 308) = 9.61$, p = .002, η^2_G = .03, such that retrospective estimates were higher with interpolated tests. There was also a significant main effect of increasing playback speed, $F(1, 308) = 4.02$, p =

.046, $\eta^2_G = .01$, such that retrospective estimates were higher with the normal speed lecture. There was no interaction, $F(1, 308) < 1$, $p = .62$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(156) = .78$, $p = .44$, $d = 0.12$.

With respect to prospective calibration, there was a significant main effect of testing, $F(1, 308) = 14.06$, $p < .001$, $\eta^2_G = .04$, such that participants were better calibrated in the lectures with interpolated tests. There was no effect of increasing playback speed, $F(1, 308) < 1$, $p = .81$, $\eta^2_G < .01$, or an interaction between testing and increasing playback speed, $F(1, 308) < 1$, $p = .68$, $\eta^2_G < .01$, on prospective calibration. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 2.77$, $p = .006$, $d = 0.44$, such that participants were better in the latter condition.

There was no significant main effect of testing, $F(1, 308) < 1$, $p = .36$, $\eta^2_G < .01$, increasing playback speed, $F(1, 308) = 2.22$, $p = .14$, $\eta^2_G = .01$, or an interaction, $F(1, 308) < 1$, $p = .94$, $\eta^2_G < .01$, on retrospective calibration. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(156) = 0.39$, $p = .70$, $d = 0.06$.

Effort

There were significant main effects of both testing, $F(1, 308) = 4.22$, $p = .04$, $\eta^2_G = .01$, and increasing playback speed, $F(1, 308) = 13.40$, $p < .001$, $\eta^2_G = .04$, on goal-driven effort, such that participants invested more effort when the lecture had interpolated tests and when it was speeded. There was no interaction, $F(1, 308) < 1$, $p = .71$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 4.31$, $p < .001$, $d = 0.69$, such that invested effort was higher in the latter condition.

There were significant main effects of both testing, $F(1, 308) = 8.97$, $p = .003$, $\eta^2_G = .03$, and increasing playback speed, $F(1, 308) = 17.21$, $p < .001$, $\eta^2_G = .05$, on data-driven effort, such that participants found the lecture to require more effort when the lecture had interpolated tests and when it was speeded. There was no interaction, $F(1, 308) < 1$, $p = .93$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 5.35$, $p < .001$, $d = 0.85$, such that required effort was higher in the latter condition.

Affect

There was no main effect of testing on liking, $F(1, 308) = 1.36$, $p = .24$, $\eta^2_G < .01$. There was a significant main effect of increasing playback speed, $F(1, 308) = 6.23$, $p = .01$, $\eta^2_G = .02$, such that liking was lower in the speeded lectures. There was no significant interaction, $F(1, 308) < 1$, $p = .77$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 2.74$, $p = .007$, $d = 0.44$, such that liking was lower in latter condition.

There was no main effect of testing on the likelihood of watching another lecture in the same format, $F(1, 308) < 1$, $p = .33$, $\eta^2_G < .01$. There was a significant main effect of increasing playback speed, $F(1, 308) = 5.15$, $p = .02$, $\eta^2_G = .02$, such that the likelihood of watching a similar lecture in the future was lower in the speeded lectures. There was no significant interaction, $F(1, 308) < 1$, $p = .98$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 2.37$, $p = .02$, $d = 0.38$, such that the likelihood of watching a similar lecture in the future was lower in the latter condition.

There was no significant main effect of testing, $F(1, 308) < 1$, $p = .86$, $\eta^2_G < .01$, increasing playback speed, $F(1, 308) < 1$, $p = .36$, $\eta^2_G < .01$, or an interaction, $F(1, 308) < 1$, $p = .51$, $\eta^2_G < .01$, on positive PANAS. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded no significant difference, $t(156) = 0.77$, $p = .44$, $d = 0.12$.

There was no main effect of testing on negative PANAS, $F(1, 308) < 1$, $p = .57$, $\eta^2_G < .01$. There was a significant main effect of increasing playback speed, $F(1, 308) = 29.73$, $p < .001$, $\eta^2_G = .09$, such that negative affect reports were higher for participants who watched the speeded lecture. There was no interaction, $F(1, 308) < 1$, $p = .79$, $\eta^2_G < .01$. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(156) = 4.58$, $p < .001$, $d = 0.73$, such that negative affect reports were higher in the latter condition.

Discussion

The results of Experiment 2 largely replicated those of Experiment 1. There was a clear beneficial effect of interpolated testing on items that had been tested (segments 1, 2, and 3) and items that had not been tested (segment 4). Unlike Experiment 1, there was an overall cost of increasing playback speed. This cost was significant for segments 1, 2, and 3 but not segment 4. As in Experiment 1, there was no interaction between interpolated testing and increasing playback speed. When we compared lectures that took an equivalent amount of time, the lecture with interpolated tests that was speeded did not significantly differ from the normal speed lecture that had no interpolated tests, though the effect was marginal. This reduction in the effect size for this particular contrast (Experiment 1: d = .51 and .49; Experiment 2: d = .28 and .30 for segments 1, 2, and 3 and segment 4 respectively) makes sense against the backdrop of a larger cost to speeding (i.e., in Experiment 1 there was no cost). Indeed, the cost of increasing playback speed in Experiment 2 provides an important reminder that increasing playback speed can impair memory for lecture material (Jacobson et al., 2018; Murphy et al., 2022). In the classroom the question is whether any such cost is worth the benefit obtained via increasing playback speed.

Moving beyond memory for lecture material, as in Experiment 1, across all of the measures here we found no interaction between increasing playback speed and the inclusion of interpolated tests. However, we did find a number of main effects. Participants' prospective performance estimates were reduced in the interpolated testing condition (this was not significant in Experiment 1) and their prospective calibration increased; their retrospective performance estimates were higher in the condition with interpolated tests, while there was no difference in retrospective calibration. With respect to affect, the inclusion of interpolated tests had no effect. While in Experiment 1, there was an increase in positive affect when interpolated tests were included, this was not the case in Experiment 2. Lastly, with respect to effort, individuals reported investing more effort in the lectures with interpolated tests and reported that they required more effort. The latter result differed from Experiment 1. The modest differences between Experiment 1 and 2 in terms of positive affect and required effort might reflect the fact that both of these measures were placed before the final test in Experiment 2.

Turning to increasing playback speed, as in Experiment 1, there was largely no effect on any metacognitive variables, though there was a significant negative effect of speed on retrospective estimates, consistent with actual performance. The lack of an effect of playback speed on prospective estimates in Experiment 2 is interesting given that there was a significant effect of increasing playback speed on memory for the lecture material. This result suggests that individuals might not always be sensitive to the negative impacts of speeding on their learning (at least when those effects are modest). Increasing playback speed did impact measures of affect. For sped-up lectures participants reported lower liking, less willingness to watch a similar lecture in the future, and more negative affect. Lastly with respect to effort, increasing playback again increased subjective effort both in terms of what learners reported investing and what learners perceived to be required.

Combined Analysis

Provided the similarity between Experiment 1 and 2, we also report a combined analysis. This was not pre-registered.

Test Performance

There was an overall significant main effect of testing on the combined accuracy of segments 1, 2, and 3, $F(1, 625) = 45.38$, $p < .001$, $\eta^2_G = .07$, and segment 4, $F(1, 625) = 33.07$, $p < .001$, $\eta^2_G = .05$, such that testing led to higher accuracy. With respect to accuracy of segments 1, 2, and 3, there was an interaction between increasing playback speed and Experiment, $F(1, 625) = 4.56$, $p = .03$, $\eta^2_G = .01$, such that the effect of increasing playback speed was smaller in Experiment 1 than Experiment 2. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference in accuracy for segments 1, 2, and 3, $t(318) = 3.52$, $p < .001$, $d = .39$, and segment 4, $t(318) = 3.55$, p < .001, $d = 0.40$, favoring the speeded condition with interpolated tests. No other effects were significant.

Metacognition

There was an overall significant main effect of testing on prospective performance estimates, $F(1, 625) = 6.19$, $p = .01$, $\eta^2_G = .01$, such that participants' estimates were higher in the no interpolated tests condition, and a main effect of Experiment, $F(1, 625) = 12.93$, $p < .001$, $\eta^2_G = .02$, such that participants provided higher estimates in Experiment 1. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(318) = 2.16$, $p = .03$, $d = 0.24$, such that prospective performance estimates were higher in the former condition. No other effects were significant.

There was an overall significant main effect of testing on retrospective performance estimates, $F(1, 625) = 28.04$, $p < .001$, $\eta^2_G = .04$, such that participant's estimates were higher in the interpolated tests condition. There was a main effect of Experiment, $F(1, 625) = 5.46$, $p = .02$, $\eta^2_G = .01$, such that participants provided higher estimates in Experiment 1 and an interaction between increasing playback speed and Experiment, $F(1, 625) = 5.46$, $p = .02$, $\eta^2_G = .01$, such that the difference in retrospective performance estimates was larger in Experiment 2, mirroring actual performance. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(318) = 3.38$, $p < .001$, $d = 0.38$, such that estimates were higher in the latter condition.

With respect to prospective calibration, there was a significant main effect of testing, $F(1, 625) = 26.46$, $p < .001$, $\eta^2_G = .04$, such that participants were better calibrated in the lectures with interpolated tests. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference on prospective calibration, $t(318) = 3.25$, $p = .001$, $d = 0.36$, such that participants were better calibrated in the latter condition. No other effects were significant and there were no significant effects on retrospective calibration.

Effort

There were overall significant main effects of both testing, $F(1, 625) = 14.42$, $p < .001$, $\eta^2_G = .02$; $F(1, 625) = 10.66$, $p = .001$, $\eta^2_G = .02$, and increasing playback speed, $F(1, 625) = 16.20$, $p < .001$, $\eta^2_G = .03$; $F(1, 625) = 19.69$, $p < .001$, $\eta^2_G = .03$, on goal-driven and data-driven effort, respectively, such that participants invested more effort and thought the lecture required more effort when the lecture had interpolated tests and when it was speeded. There was a main effect of Experiment, $F(1, 625) = 5.86$, $p = .01$, $\eta^2_G = .01$, such that goal-driven effort reports were higher in Experiment 2. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(318) = 5.54$, $p < .001$, $d = 0.62$; $t(318) = 5.55$, $p < .001$, $d = 0.62$, such that both invested effort and required effect, respectively, were higher in the increased speed lecture with interpolated tests. No other effects were significant.

Affect

There was an overall significant main effect of increasing playback speed on liking, $F(1, 625) = 13.16$, $p < .001$, $\eta^2_G = .02$, and the likelihood of watching a similar lecture, $F(1, 625) = 18.72$, $p < .001$, $\eta^2_G = .03$, such that liking and the likelihood of watching a similar lecture was lower in the speeded lectures. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference for liking, $t(318) = 2.59$, $p = .01$, $d = 0.29$, and the likelihood of watching a similar lecture, $t(318) = 2.37$, $p = .02$, $d = 0.38$, such that they were both lower in the increased speed with interpolated tests condition. No other effects were significant.

There was a significant main effect of Experiment on positive PANAS, $F(1, 625) = 4.52$, $p = .03$, $\eta^2_G = .01$, such that positive affect scores were higher in Experiment 2. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(318) = 2.30$, $p = .03$, $d = 0.26$, such that positive affect scores were higher in the latter condition. No other effects were significant.

There was overall a significant main effect of increasing playback speed on negative PANAS, $F(1, 625) = 30.50$, $p < .001$, $\eta^2_G = .05$, such that negative affect scores were higher for participants who watched the speeded lecture. There was a main effect of Experiment, $F(1, 625) = 6.05$, $p = .01$, $\eta^2_G = .01$, such that negative affect scores were higher in Experiment 1. The direct comparison between the normal speed lecture with no interpolated tests and the increased speed lecture with interpolated tests yielded a significant difference, $t(318) = 4.51$, $p < .001$, $d = 0.50$, such that negative affect scores were higher in the latter condition.

General Discussion

The present investigation began with a basic problem: testing is now widely regarded as an effective pedagogical strategy (Agarwal et al., 2021; Dunlosky et al., 2013; Rowland, 2014) but that its implementation requires additional instructional time which itself carries an associated cost. We examined one solution to this problem in the context of the use of interpolated tests in recorded lectures (Haagsman et al., 2020; Jing et al., 2016; Lavigne & Risko, 2018; Szpunar et al., 2013; 2014; van der Meij & Bockmann, 2021; Yang et al., 2020). Namely, we examined whether increasing the playback speed of the lecture could be used to reduce the time cost without compromising the benefit of interpolated testing in terms of memory for lecture material. By examining the conjoint effects of interpolated testing and increasing playback speed we could provide insight into a potential solution to a pedagogical problem. Beyond memory for lecture material, we also examined the influence of interpolated testing, increasing playback speed, and their combination on individual's metacognitions, subjective effort, and affect in order to provide further insight into these strategies.

Turning first to memory for lecture material. The results were clear. Overall, interpolated testing benefitted memory for the lecture material. This was true whether the items tested had been tested during the lecture (i.e., segments 1-3) or not (i.e., segment 4). Thus, interpolated testing produced both a backward and forward testing effect. Increasing playback speed did not impact memory for lecture material in Experiment 1 but did in Experiment 2. Critically, there was no interaction between interpolated testing and increasing playback speed, that is, the benefit of interpolated testing was the same magnitude whether the lecture was played at its normal speed or had its playback speed increased. This result suggests, within the confines of the present study at least, that increasing playback speed represents a viable method for reducing the instructional time costs associated with using interpolated testing as a pedagogical strategy. Increasing playback speed (a modest amount) did not compromise the benefit of interpolated testing. The caveat noted above (i.e., within the confines of the present study) is important to keep in mind when considering adopting these approaches. For example, this study consisted of only a single study session (e.g., a single lecture) and as such does not capture potentially cumulative effects that might emerge in courses wherein individuals view many recorded lectures.

While the combined effects of interpolated tests and increasing playback speed were additive, both did have an impact on load (operationalized here via subjective effort; Koriat & Nussinson, 2009). That is, adding interpolated tests and increasing playback speed increased the self-reported required effort and invested effort (except for required effort as a function of adding interpolated tests in Experiment 1). As with memory for lecture material, there was no interaction between adding interpolated tests and increasing playback speed. With respect to adding interpolated tests, this result appears inconsistent with Szpunar et al. (2013) who reported that individuals in their interpolated test condition reported the experience of learning the lecture to be less "mentally taxing" than individuals in their restudy and non-tested group. This difference might reflect the different wording of the items. Another important difference is that here the "control" represented a lecture without interpolated tests whereas in Szpunar et al. (2013) the "control" featured interpolated tests with answers (thus removing the testing but maintaining re-exposure) or solving math problems in the non-tested group. These different controls represent the different goals of each of the projects. The controls used in Szpunar et al. (2013) provided a better opportunity to draw inferences about the influence of interpolated tests (which was their goal). Here, the interest was to have the results directly inform instructional design decisions which would involve the choice between a normal lecture and adding interpolated tests, increasing playback speed or both (not adding interpolated tests or adding interpolated tests with answers or interpolating math questions). As such, we think it safe to at least preliminarily conclude that the addition of interpolated testing to a recorded lecture will tend to increase the learner's load (in terms of required and invested effort) relative to a recorded lecture without interpolated tests.

In Experiment 1, an increase in effort was not accompanied by a decrease in memory for the lecture material. One straightforward interpretation of this result is that load imposed by the normal speed lecture is such that there is enough capacity to increase the investment of resources without compromising performance. For example, in the context of cognitive load theory, the lecture might be low in intrinsic complexity. As noted in the Introduction, there exist numerous demonstrations that one can increase playback speed without causing a decrement in performance

(Murphy et al., 2022; Nagahama & Morita, 2017, 2018; Wilson et al., 2018). The idea that in these situations the normal speed lecture's intrinsic characteristics permit the increases in load induced by increasing playback speed without leading to the overloading of the learner represents a reasonable account of this data. A different (though not necessarily competing) explanation for why increased effort was not associated with decreased performance is that the increase in load did impair learning but the test was not sensitive enough to pick it up. Future research focused on when (not simply whether) increasing playback speed increases effort and impairs performance would be valuable.

With respect to metacognition, as with memory for the lecture material, participants' performance estimates prior to and after the test and their calibration showed no evidence of an interaction between interpolated testing and increasing playback speed. Nevertheless, there were several patterns worth noting. First, the performance estimates prior to the test replicated Szpunar et al. (2014). Specifically, adding interpolated tests led to better calibration in the judgment prior to the test (i.e., the absolute accuracy of their estimates were closer to their actual performance). After the test, participants who had interpolated tests did feel more confident in their performance than individuals who had not had interpolated tests. This feeling was accurate in the sense that the former did perform better. The better calibration in the interpolated testing group prior to the test also disapeared in the estimate after the test. Taking the test improved calibration in the no interpolated testing groups but had little effect on the group with interpolated testing. These results are generally consistent with the prediction derived in the Introduction wherein being tested aids participants metacognitively. That is, the group who had been tested prior to the cumulative test had better calibration than the group who had not, but once both groups had been exposed to a test calibration was similar.

Turning to affect, adding interpolated tests had limited influence on affective variables. The one exception was an increase in positive affect in Experiment 1 in the interpolated testing condition. This was not replicated in Experiment 2 where the PANAS occurred before the final test, instead of after as in Experiment 1. This leaves open the possibility that the result in Experiment 1 reflects performance on the test (where the interpolated test group did better). That said, a genuine effect on positive affect could also be interpreted as the tests increasing activation. Positive PANAS items included emotions like "alert," "attentive," and "active," all of which would likely increase if individuals were investing more effort as indicated by the effort measures above. While adding interpolated tests had a limited effect on affect, increasing playback speed had a pronounced effect on affect. Increasing playback speed reduced liking, reduced the likelihood that an individual would watch a similarly formatted lecture in the future, and increased negative affect. Thus, increasing playback speed seems to create a less positive learning experience. Note that this was true in both Experiment 1, where there was no influence of increasing playback speed on test performance, and in Experiment 2, where there was a negative influence of increasing playback speed on test performance. Returning to the notion that increasing playback speed might represent a strategy compensating for added instructional time associated with testing, the robust negative influence of speeding should give the instructional designer some pause about using this strategy. That said, when considering the affective impact of increasing playback speed, it is important to note that we had no measure of the experience our participants had with increasing playback speed. In addition, our sample was not restricted to current post-secondary students, a population who might be expected to have extensive experience with this strategy, thus raising the possibility that the negative experience we found was at least partly due to a lack of familiarity with consuming material at an increased playback speed. Future work including this variable would be valuable.

Conclusion

Making testing an integral part of instructional design appears to represent an evidence-based method for improving learning. Doing so, however, brings forth challenges with respect to implementation. Here we examined one such challenge – the increased instructional time required when interpolated tests are added to a lecture – and one solution to that challenge – increasing the playback speed of the lecture. Critically, increasing the playback speed sufficient to counteract the increased time required for interpolated testing did not reduce the benefit of interpolated testing. That is, this strategy was effective at least in terms of performance on a test designed to measure memory for the lecture material. That said, increasing playback speed was associated with costs that need to be taken into consideration. Beyond the conjoint effects of adding interpolated tests and increasing playback speed, the present investigation has also provided added insight into the use of each of these strategies and raised important new questions that need to be addressed.

Footnotes

1. Prior to the reported Experiment 1 we conducted a similar experiment with a smaller sample size using only a subset of the conditions including (1) Increased speed with interpolated tests (2) Standard speed with interpolated tests and (3) Standard speed with no interpolated tests. There was an effect of condition on segments 1, 2, and 3 post-test performance, wherein the increased speed with tests condition was located between the other two conditions. This was ambiguous to interpret without an increased speed no tests condition. There was no effect of condition on segment 4 performance, estimates after, calibration after, goal driven effort, or positive or negative PANAS. Performance estimates prior to the test were lower, calibration higher, liking and likelihood of watching a similar lecture in the future lower, in the Increased speed with tests condition relative to the Standard speed with no tests condition. The latter variables did not differ across the Increased speed with tests condition and the normal speed with tests condition except for a reduced likelihood of watching a similar lecture in the future in the Increased speed with tests condition. Lastly, participants reported that the increased speed with tests condition required more effort than the normal speed with tests condition. Provided the lack of the full factorial design we did not include a full report of this experiment.
2. We thank Karl Szpunar for sharing the materials.

References

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, *24*(3), 437–448. https://doi.org/10.1007/s10648-012-9210-2

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, *33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Ahn, D., & Chan, J. C. K. (2022). Does testing enhance new learning because it insulates against proactive interference? *Memory & Cognition*, 50, 1664-1682. https://doi.org/10.3758/s13421-022-01273-7

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

Barron, A. E. (2013). Auditory instruction. In D. H. Jonassen (Ed.), Handbook of Research for Educational Communications and Technology (pp. 949–978). Mahwah, NJ: Lawrence Erlbaum.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1552.

Dunlosky, J., & Rawson, K. A. (2019). How cognitive psychology can inform evidence-based education reform: An overview of the cambridge handbook of cognition and education. In

J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 1–14). Cambridge University Press. https://doi.org/10.1017/9781108235631.001

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in Psychology*, *4*, 619. https://doi.org/10.3389/fpsyg.2013.00619

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, *45*(3), 349–359. https://doi.org/10.1518/hfes.45.3.349.27253

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 41–50. https://doi.org/10.1145/2556325.2566239

Haagsman, M. E., Scager, K., Boonstra, J., & Koster, M. C. (2020). Pop-up questions within educational videos: Effects on students' learning. *Journal of Science Education and Technology*, *29*(6), 713–724. https://doi.org/10.1007/s10956-020-09847-3

Jacobson, L. A., Geist, M., & Mahone, E. M. (2018). Sluggish cognitive tempo, processing speed, and internalizing symptoms: The moderating effect of age. *Journal of Abnormal Child Psychology*, *46*(1), 127–135. https://doi.org/10.1007/s10802-017-0281-x

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*(3), 305–318. https://doi.org/10.1037/xap0000087

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need?. *Educational Psychology Review*, *23*(1), 1-19.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Lavigne, E., & Risko, E. F. (2018). Optimizing the use of interpolated tests: The influence of interpolated test lag. *Scholarship of Teaching and Learning in Psychology*, *4*(4), 211–221. https://doi.org/10.1037/stl0000118

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012.

Lab, C. (2023). The Effect of Lecture Speeding and Interpolated Testing - Experiment 2. Retrieved from osf.io/254pt

Liu, J., Bianchi, L., & Lab, C. (2023). The Effect of Lecture Speeding and Interpolated Testing - Experiment 1. Retrieved from osf.io/hmjta

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, *37*(6), 661–670. https://doi.org/10.1016/S0028-3932(98)00127-4

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801_6

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Murphy, D. H., Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology*, *36*(1), 69–82. https://doi.org/10.1002/acp.3899

Nagahama, T., & Morita, Y. (2017). Effect analysis of playback speed for lecture video including instructor images. *International Journal for Educational Media and Technology*, *11,* 50-58.

Nagahama, T., & Morita, Y. (2018). Analysis of learning effectiveness with high-speed visual content. *Educational Technology Research and Development, 40,* 85-97.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. https://doi.org/10.1007/BF02504795

Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive Load Theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*(2), 115–121. https://doi.org/10.1007/s10648-010-9133-8

Pastore, R., & Ritzhaupt, A. D. (2015). Using time-compression to make multimedia learning more efficient: Current research and practice. *TechTrends*, *59*(2), 66–74. https://doi.org/10.1007/s11528-015-0841-2

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in Cognitive Load Theory. *Educational Psychology Review*, *31*(2), 339–359. https://doi.org/10.1007/s10648-019-09473-5

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*(2), 234–242. https://doi.org/10.1002/acp.1814

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432. https://doi.org/10.1037/a0037559

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Seli, P., Wammes, J. D., Risko, E. F., & Smilek, D. (2016). On the relation between motivation and retention in educational contexts: The role of intentional and unintentional mind wandering. *Psychonomic Bulletin & Review*, *23*(4), 1280–1287. https://doi.org/10.3758/s13423-015-0979-0

Sweller, J. (2011). Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, *3*(3), 161–164. https://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*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(6), 1392–1399. https://doi.org/10.1037/a0013082

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117 van der Meij, H., & Böckmann, L. (2021). Effects of embedded questions in recorded lectures. *Journal of Computing in Higher Education*, *33*(1), 235–254. https://doi.org/10.1007/s12528-020-09263-x

van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*(2), 265–289. https://doi.org/10.1007/s10648-015-9297-3

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x.

Van Kessel, K., Ashburner, M., Risko, E. F. (in press). Dissociations between data-driven and goal-driven effort reports: Performance, metacognition, and affect. Quarterly Journal of Experimental Psychology.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063

Welhaf, M. S., Phillips, N. E., Smeekens, B. A., Miyake, A., & Kane, M. J. (2022). Interpolated testing and content pretesting as interventions to reduce task-unrelated thoughts during a video lecture. *Cognitive Research: Principles and Implications*, *7*(1), 26. https://doi.org/10.1186/s41235-022-00372-y

Wilson, K. E., Martin, L., Smilek, D., & Risko, E. F. (2018). The benefits and costs of speed watching video lectures. *Scholarship of Teaching and Learning in Psychology, 4*(4), 243–257. https://doi.org/10.1037/stl0000127

Yang, C., Chew, S.-J., Sun, B., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning. *Journal of Educational Psychology*, *111*, 809–826. https://doi.org/10.1037/edu0000320

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *NPJ Science of Learning*, *3*(1), 1–9. https://doi.org/10.1038/s41539-018-0024-y

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*(4), 724–738. https://doi.org/10.1037/xap0000278

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2021). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1127–1143. https://doi.org/10.1037/xlm0001021