Journal of Applied Research in Memory and Cognition 3 (2014) 161–164

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education

Karl K. Szpunar*, Helen G. Jing, Daniel L. Schacter

Harvard University, Department of Psychology, United States

ARTICLE INFO

Article history:
Received 3 December 2013
Received in revised form 20 February 2014
Accepted 25 February 2014
Available online 3 March 2014

Keywords:
Interpolated testing
Online learning
Video-recorded lectures
Judgments of learning

ABSTRACT

The video-recorded lecture represents a central feature of most online learning platforms. Nonetheless, little is known about how to best structure video-recorded lectures in order to optimize learning. Here, we focused on the tendency for high school and college students to be overconfident in their learning from video-recorded modules, and demonstrated that testing could be used to effectively improve the calibration between predicted and actual performance. Notably, interpolating a lecture with repeated tests helped to boost actual performance to the level of predicted performance, whereas a single test following the lecture served to lower unrealistic judgments of learning. The value of improving performance to match predictions of learning and other avenues for future research regarding meta-comprehension of video-recorded lectures is discussed.

1. Introduction

The video-recorded lecture represents a central feature of most online learning platforms (Breslow et al., 2013). Nonetheless, little remains known about what obstacles students might encounter when learning from video-recorded lectures or how those obstacles might be overcome. Here, we focus on how well students think they will perform on a later assessment associated with learning from a video-recorded lecture. Considerable research has indicated that students overestimate their ability to assess later performance associated with learning from video-recorded modules (Choi & Johnson, 2005; Salomon, 1984; for a recent review, see Means, Toyama, Murphy, Bakia, & Jones, 2010). Importantly, overconfidence in later performance can have a negative impact on long-term retention. For instance, students making overconfident judgments of learning have been shown to cut short subsequent opportunities for re-study (Dunlosky & Rawson, 2012; see also Bol & Hacker, 2012).

Further complicating matters, students tend to hold stable persistent beliefs about how well they learn in traditional educational settings (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Schraw, Potenza, & Nebelsick-Gullet, 1993). For instance, various studies have demonstrated that it can be difficult to alter global judgments of learning that are based, at least partly, on lecture content (e.g., judgments of learning for mid-term exams; for a recent review, see Hacker, Bol, & Keener, 2008). Given that students overestimate how well they will perform on subsequent assessments associated with video-recorded materials and that this metacognitive error may be difficult to correct, what can be done to improve calibration between predicted and actual performance in online learning environments?

One approach may be to seek interventions that re-structure lectures in a manner that can boost actual performance to the level of predicted performance. Along these lines, considerable research has demonstrated that the act of retrieving information from memory can serve to boost learning and retention in educational settings (for relevant reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006). Indeed, we recently demonstrated that interpolating a video-recorded lecture with brief memory tests served to substantially enhance learning (Szpunar, Khan, & Schacter, 2013). In the present study, we sought to examine whether interpolated testing during a lecture would elevate actual performance to the level of predicted performance.

The study involved three groups of high school students who learned from a statistics lecture. The use of video-recorded lectures in the context of online learning is quickly becoming a popular method of delivering educational content with high school populations (Picciano, Seaman, Shea, & Swan, 2012). Moreover, statistics is commonly perceived as being especially difficult to master (Gal & Ginsburg, 1994), and so any indication of overconfidence in learning from a statistics lecture would further highlight the robust

* Corresponding author at: Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA. Tel.: +1 617 495 9031.
E-mail address: szpunar@wjh.harvard.edu (K.K. Szpunar).

http://dx.doi.org/10.1016/j.jarmac.2014.02.001
2211-3681/

nature of overconfidence in learning from video-recorded materials.

One group of high school students learned the lecture in the presence of interpolated tests, whereas another group of high school students learned the lecture in the absence of interpolated tests. Although we did not expect global judgments of learning to differ between the two groups, we predicted that interpolated testing should nonetheless boost final test scores in a manner that would better align predicted and actual performance. Hence, students in the interpolated group should appear less overconfident and better calibrated. An alternative hypothesis is that interpolated testing could boost both actual and predicted performance and hence not improve calibration. Specifically, it is possible that students may find the act of answering questions during the lecture easy, which could serve to elevate predictions of future performance (cf. Schunk, 1991). To test the generalizability of the effects of interpolated testing on actual and predicted performance, we also carried out a partial re-analysis of an existing dataset that involved college students learning from the same statistics lecture under conditions of interpolated and non-interpolated testing.

Finally, we included a third group of high school students that was also afforded the opportunity to answer questions during initial learning of the lecture, but only after the final portion of the lecture. We have previously shown that students experience considerable difficulty answering questions after an extended period of study that does not involve interpolated testing (Szpunar, McDermott, & Roediger, 2008; Szpunar et al., 2013), and sought to assess whether this salient experience with difficult-to-answer questions would help students to lower unrealistic judgments of learning.

2. Method

2.1. Subjects

Fifty-four high-school students (ages 16–18 years) attending Harvard University's summer school program participated in the study. High-school students were recruited because they had little-to-no prior experience with statistics. Students provided informed written consent, obtained parental consent if they were younger than 18 years, and were randomly assigned to one of three experimental groups.

2.2. Study materials

An introductory statistics lecture was used in the experiment (Statistics 104, Department of Economics, Harvard University). The 21-min video covered basic introductory concepts in statistics (e.g., outlining the relation between a sample and population) that did not require past experience with statistics. The video was divided into four 5-min segments using iMovie software (Apple).

2.3. Design and procedure

Students took part in a 1-h learning session. Students were told that the video-recorded lecture would be divided into four segments of equal length. Further, students were told that they would complete a number of tasks between each segment. Initially, students were informed that they would complete 1 min of math problems after each segment of the lecture that was unrelated to the content of the lecture (six problems were presented and students were given 10 s to answer each problem; e.g., $12 \times 7 = ?$). Moreover, students were told that either two more minutes of math problems (12 problems; 10 s per problem) or a 2-min quiz about the most recent segment of the lecture (six questions; 20 s per question; e.g., What is the relation between a sample and population?) would follow the first minute of math problems (following each segment). Note that the test questions were brief short answer questions that tapped memory for information explicitly presented in the lecture. Importantly, students were informed that a computer program would randomly determine the occurrence of the quizzes, such that students might experience 0–4 quizzes during the lecture. For example, students could be quizzed after each lecture segment, after none of the segments, or anywhere in between. Finally, students were told that regardless of the frequency of testing during the lecture that there would be a final cumulative test that would test their knowledge about the entire lecture. In reality, one-third of the students received tests after all four lecture segments (4-test group), one-third of the students received a test following the fourth and final segment of the lecture (1-test group), and one-third of the students did not receive any tests during the lecture (0-test group). After the lecture was complete, students were given a 5-min break during which they played an online computer game (Tetris). After the break, students were asked to predict, on a scale of 0–100%, how well they thought that they would perform on the final cumulative test. The final cumulative test included the same 24 questions that were presented to students in the interpolated group, and students were allowed to complete the final test in a self-paced manner. Note that students were not given any indication about the types of questions that they would receive on the final test. The lecture, math questions, and quiz questions were presented on a computer screen using E-Prime 2.0 software on a Dell desktop computer, and responses were made using a keyboard.

Finally, we set out to assess whether our previous demonstration that interpolated tests helped students to avoid mind wandering and engage in note taking (Szpunar et al., 2013) would extend to a population of high-school students. In order to do so, our experimental design also incorporated the following features. First, students were told that a visual cue indicated by the phrase "Mind wandering? Yes/No" would appear on the computer screen at some random points during the lecture, and that whenever they saw this cue that they should respond on a sheet of paper by writing the word 'yes' or 'no'. The visual cue remained on the screen for 5 s as the lecture continued, and was accompanied by an auditory cue (i.e., a bell) that sounded during the first of the 5 s to ensure that students noticed the cue. Four mind wandering probe sequences were used in the study. For each sequence, the mind wandering probe occurred at a randomly determined time point during each segment that was at least 30 s into the segment and 30 s before the segment was complete. For instance, one sequence involved probes that occurred 96 s, 285 s, 201 s, and 155 s into the first, second, third, and fourth segments, respectively. The presentation of these mind wandering probe sequences was counterbalanced across students. Second, students were provided with the lecture slides associated with the lecture, and instructed that they could use the slides in any way that they thought might help them learn from the lecture. Upon the completion of the lecture, we retrieved each student's lecture slides. As a rough measure of student engagement, we checked to see for what proportion of slides students took additional notes, and whether interpolated testing influenced note-taking behavior. Note taking was defined in a manner such that both additional notes associated with lecture content and emphasis given to existing notes (e.g., circling or underlining key lecture points) were counted as additional notes. However, markings unrelated to lecture content (e.g., doodles) were excluded from the analysis.

3. Results

3.1. Initial test

In order to determine that students in the 1-test group were in fact making predictions of final test performance following an