Effects of embedded questions in recorded lectures

Hans van der Meij[1] · **Linn Böckmann[2]**

Abstract

Online video-recorded lectures have become an increasingly more important means for student learning (e.g., in flipped classrooms). However, getting students to process these lectures sufficiently to come to class well-prepared is a challenge for educators. This paper investigates the effectiveness of open-ended embedded questions for accomplishing that. An experiment compared a video-recorded lecture presented online with and without such questions. No feedback was given on responses to the questions. University students ($N = 40$) viewed the lecture, responded to a questionnaire on self-efficacy and usability, and completed a knowledge test. User logs revealed that the students engaged significantly more with the embedded questions lecture. Engagement was not related to knowledge test results, however. Uniformly high appraisals were given for self-efficacy, usefulness, ease of use and satisfaction. Mean test scores were significantly higher for the embedded questions condition. It is concluded that open-ended embedded questions without feedback can increase the effectiveness of online video-recorded lectures as learning resources.

Keywords Quizzing · Embedded questions · Video-recorded lectures · Self-efficacy · Usability

Flipped classrooms (FCs) are rapidly gaining popularity (Fredriksen 2020; Karabulut-Igu et al. 2018; Turan and Akdag-Cimen 2020). In FCs, the delivery of instructional content takes place before class and outside the classroom. Students often need to prepare by watching a video-recorded lecture online at home, so that classroom time can be spent on student-centered learning activities (Lage et al. 2000). Because generally positive effects for FCs have been found (e.g., Akçayır and Akçayır 2018; Bond 2020; Strelan et al. 2020), systematic research is being conducted on the main contributing factors.

✉ Hans van der Meij
H.vanderMeij@utwente.nl

[1] Faculty of Behavioural, Management and Social Sciences, Department of Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

[2] Lerntherapeutische Praxis Dülmen, Soester Straße 66, 48155 Münster, Germany

Springer

The video-recorded lecture is such a key factor, and much of the research has focused on what makes the video-recorded lecture (more) effective (e.g., Lin and Chen 2019; Toftness et al. 2018a; Zhang et al. 2006). One design feature that is studied is quizzing in which a video-recorded lecture is complemented with questions to stimulate more active or deeper lecture processing (e.g., Christiansen et al. 2017; Cummins et al. 2016; Kovacs 2016). The present study investigates the impact of an understudied variant of quizzing, namely, the inclusion of open-ended, embedded questions.

Research on adjunct questions with texts (e.g., Andre 1981; Ozgungor and Guthrie 2004; Uner and Roediger 2018) suggests that the inclusion of embedded questions is likely to enhance the effectiveness of recorded lectures. Such questions can stimulate students' retrieval practice and can help them realize that they did not comprehend a message, or cannot remember key facts. This should prompt a mental review or replay of the video-recorded lecture until that goal is achieved. Empirical research on embedded questions in video-recorded lectures has already shown promising results (e.g., Cummins et al. 2016; Leisner et al. 2020; Rice et al. 2019), but important design variations and effects of such questions appear to have been understudied thus far (e.g., Brinton et al. 2016; Haagsman et al. 2020; Ketsman et al. 2018; Kovacs 2016).

One design variation that merits systematic research is question format. Many studies on embedded questions in live lectures involve Audience Response Systems that generally present multiple-choice questions (e.g., Buil et al. 2016; Crandall et al. 2019; Hunsu et al. 2016; Khan et al. 2019; Pan et al. 2019; Shapiro and Gordon 2012; Shapiro et al. 2017). While it might be tempting to also use such questions in video-recorded lectures, these may not be the most effective kinds of questions to ask. If the aim is to stimulate student recall of the learning material, open-ended questions are probably more suitable than multiple-choice questions, which rely more on recognition (Rawson and Dunlosky 2012).

As yet, though, there have been few empirical studies involving open-ended embedded questions in video-recorded lectures (Cummins et al. 2016; Szpunar et al. 2013, 2014; Thomas et al. 2018). Three of the four studies found did not support a firm conclusion that open-ended questions result in better learning than no questions because these questions were blended with other stimuli for active video processing. That is, in the study of Cummins et al. (2016) open-ended and multiple-choice questions were mixed and in the studies of Szpunar et al. (2013, 2014) there were practice items along with open-ended questions. To our knowledge, the only investigation comparing open-ended questions with a non-question condition is a recent study by Thomas and colleagues (Thomas et al. 2018). That study found a positive effect of these questions on learning. In the present study, the intervention consists of asking open-ended embedded questions within a video-recorded lecture, which is compared to the lecture with no questions included.

Another design issue that should probably be further investigated is feedback. Empirical research generally shows that the presence of feedback enhances learning (Fiorella and Mayer 2018; Shute 2008). However, feedback (in the form of identification and possibly explanation of the correct answer) can pre-empt the students' active processing of the lecture that should be promoted by the presence

of questions. That is, the adjunct questions research suggests that the presence of feedback induces students to invest less effort in responding to the quiz questions and hence reduces learning (e.g., Hamaker 1986; Roelle et al. 2017). In the present study, there is no feedback on the responses to quiz questions. This distinguishes the present study from the research by Thomas and colleagues (Thomas et al. 2018), where feedback was given.

As related above, the effort to improve processing of instructional content presented in video-recorded lectures by including embedded questions is connected to the long-standing tradition of providing adjunct questions with texts (Hamilton 1985) and ties in as well with more recent research on quizzing with lectures (Brink 2013; Shapiro et al. 2017). These lines of research are described next. First, there is a discussion on the main kinds of dependent variables that are investigated in the present study. Thereafter, a detailed account is given on why open-ended embedded questions may be particularly well-suited to enhance learning from video-recorded lectures.

Adjunct questions and quizzing

Adjunct questions are questions added to an instructional text to enhance what is learned from that text (Rothkopf 1970). Meta-analyses have reported robust effects of adjunct questions on learning (Anderson and Biddle 1975; Hamaker 1986; Hamilton 1985). Texts with adjunct questions generally yield higher test outcomes than texts without such questions. Quizzing is an emergent educational trend that is reminiscent of the adjunct questions method. Quizzing is an instructional approach in which questions are included in live or video-recorded lectures to increase their effectiveness. The literature has indicated that quizzing can have a positive effect on engagement (e.g., Cummins et al. 2016; Mayer et al. 2009), appreciation and motivation (e.g., Buil et al. 2016; Zhu 2008) and learning (e.g., McDaniel et al. 2013; Shapiro et al. 2017).

Research generally shows that quizzing promotes active student *engagement* (e.g., Draper and Brown 2004; Khanna 2015; Shapiro 2009; Trees and Jackson 2007; Zhu 2008). The added presence of quizzing in live lectures has almost invariably been found to increase classroom attendance and students' active participation during class (e.g., Caldwell 2007; Khan et al. 2019; van Daele et al. 2017; Wang 2020). Quizzing in video-recorded lectures has likewise often yielded desirable engagement related outcomes such as lower in-video dropout (Vural 2013) and persistence in processing the quiz questions (Kovacs 2016). The present study focused on processing time as measure of engagement as that can play a key role in learning (e.g., Rice et al. 2019; Shinaberger 2017). The hypothesis is tested that engagement is higher for the embedded-question videos, just as has been found in other studies (e.g., Kovacs 2016; Vural 2013). In addition, the study explores whether video engagement is related to learning.

Comparisons between quizzing and non-quizzing conditions in live and video-recorded lectures have further shown that quizzing often results in higher *appreciation* and stronger *motivation* (e.g., Buil et al. 2016; Hunsu et al. 2016). For instance,

research generally indicates that students favorably appreciate the usability of video-recorded lectures (e.g., Baker et al. 2018; O'Callaghan et al. 2017; Spanjers et al. 2015). Research has not yet investigated whether such appraisals are affected by the presence of open-ended embedded questions. The present study explores this effect for the three usability measures proposed in the (extended) Technology Acceptance Model (Davis 1989), namely, usefulness, ease of use and satisfaction (Davis 1989; Davis et al. 1989; Joo et al. 2014). A recent meta-analysis on blended learning showed that quizzing was an important positive moderator for satisfaction (Spanjers et al. 2015). Accordingly, the experimental condition was expected to yield higher appraisals for this construct. No specific hypothesis was tested for the two other usability perceptions.

A motivational characteristic that is likely to influence students' willingness to study video-recorded lectures is self-efficacy which is a person's belief in the capacity to organize and execute the actions necessary to manage particular task outcomes (Bandura 1997). Self-efficacy has been found to be a predictor of future persistence and effort expenditure in comparable settings (Bandura 2012; Bandura and Locke 2003). Also, a recent meta-analysis on clickers—a technology that allows teachers to pose questions and to process the student responses during a live lecture—found that the largest non-cognitive effect of their presence was their contribution to self-efficacy development (Hunsu et al. 2016). This suggests that the added presence of quizzing (even when it involves multiple-choice questions) is likely to positively affect self-efficacy. Similarly, a recent experiment found that embedded questions in video-recorded lectures significantly increased self-efficacy (Tweissi 2016). The questions are likely to increase engagement with the lecture and support students' confidence in their capacity to comprehend the message that is conveyed. Accordingly, the present study tests the hypothesis that open-ended embedded questions enhance self-efficacy.

Finally, quizzing in live and video-recorded lectures has generally been found to increase *learning* (Gier and Kreiner 2009; Lawson et al. 2006; Morling et al. 2008; Vural 2013). The literature has offered two main explanations for this effect on knowledge development.

One account is that questions encourage active processing (Mayer et al. 2009). The questions may change a more passive reception of knowledge during a lecture into a more active knowledge construction mode. That is, they can stimulate students to be more selective in the information they attend to (Mayer et al. 2009). In addition, they may induce students to (re)structure the information to make it more comprehensible, leading to development of a schema or mental model of the lecture content (see Jing et al. 2016). Finally, the presence of questions may activate prior knowledge that is connected to the new information (see Carpenter 2011). The questions then serve an integrative role. Students connect new with existing knowledge, relating lecture content to what they already know on a topic.

Another explanation comes from the testing effect (e.g., McDaniel et al. 2011, 2013). This is the finding that students are better at remembering previously presented information on which they have been tested than they are at remembering untested information. The testing effect is ascribed to retrieval practices (McDermott et al. 2014). That is, questions may stimulate students to recall or reconstruct

information that addresses the quiz question. This active retrieval of lecture content more positively affects learning than other, more passive strategies such as summarizing or note-taking.

Research also shows that there can be important moderating factors such as context, placement and question format (e.g., Khanna 2015; Mayer et al. 2009; McDaniel et al. 2012; Toftness et al. 2018b). Many empirical studies on quizzing have been conducted in ecologically valid settings, namely, actual classrooms (e.g., Barr 2017; Shapiro et al. 2017). The studies have often included intact classes and involved existing courses that ran for weeks or even months (e.g., Batchelor 2015; Brink 2013; Shapiro et al. 2017). In addition, these studies have involved questions before, during and/or after the lectures (e.g., Carpenter et al. 2018; Khanna 2015; Shapiro and Gordon 2013). Furthermore, the questions that were asked included multiple-choice, short answer open-ended questions and combinations of the two (e.g., Mayer et al. 2009; McDaniel et al. 2012). These factors make it hard to draw firm conclusions about the effectiveness of specific quizzing arrangements (e.g., Mayer et al. 2009; Papadopoulos et al. 2018).

The present study is set up as a controlled true experiment involving a video-recorded lecture, in which only the presence of questions varies between conditions. The placement of the questions vis-à-vis the video-recorded lecture is important. Research shows that pre-questions posed before a lecture have limited effect on learning (e.g., Carpenter et al. 2018; Toftness et al. 2018b), and that embedded questions are more effective than post-questions asked after the lecture has been completed (e.g., Rice et al. 2019; Szpunar et al. 2014). The experiment therefore investigates embedded questions. These questions appeared automatically after each segment of a lecture.

Moreover, the study investigates the effectiveness of open-ended embedded questions. Multiple-choice is the most typical question type in video-recorded lectures (e.g., Garcia-Rodicio 2015; Jolley et al. 2016; Vural 2013); open-ended questions are rarely used (e.g., Szpunar et al. 2014; Thomas et al. 2018). The limited usage of open-ended questions appears at odds with their relative effectiveness. Empirical research suggests that open-ended questions may be more effective than multiple-choice questions for learning (e.g., Butler and Roediger 2007; McDaniel et al. 2007). For instance, Butler and Roediger (2007) investigated three learning techniques for processing the material presented in a lecture: studying a summary, taking a multiple-choice test or taking a short answer test. The findings revealed that the short answer test improved recall the most. An explanation for this effect was that answering open-ended questions requires students to engage in more taxing information retrieval attempts than answering multiple-choice questions, which hinges on recognizing the right answer among a number of alternatives. More generally, this research suggests that quiz questions that involve retrieval rather than recognition increase learning more (see Rawson and Dunlosky 2012).

As mentioned earlier, only a few controlled studies have investigated effects of embedded open-ended questions on learning from video-recorded lectures (Cummins et al. 2016; Szpunar et al. 2013, 2014; Thomas et al. 2018). In these studies, the experiment used a combination of open-ended questions with multiple choice questions, practice items, or feedback. The present study had no other support for

learning than the open-ended questions. The tested hypothesis is that open-ended embedded questions increase knowledge of lecture content.

In short, the present study compared an experimental condition in which open-ended questions were asked in a video-recorded lecture with a control condition without such questions. Three research questions were investigated:

RQ1: What is the effect of condition on video engagement?
RQ2: What is the effect of condition on technology acceptance and self-efficacy?
RQ3: What is the effect of condition on knowledge development?

Method

Participants

Forty social science students from the University of Twente volunteered to participate in the study. All students were fluent German speakers. The study included 10 male and 30 female students, with a mean age of 21.6 years ($SD = 1.96$). Students were randomly assigned to the control or experimental condition. Students received one credit point and cash payment of €7.50 for participation. Approval for the study was obtained from the Ethical Committee of the University. All instructional materials were in German.

Instructional instruments

The recorded lecture was drawn from YouTube (inCITI Singen 2015). It presented a public talk by Prof. Dr. Manfred Spitzer. The talk's setting resembled a conference keynote speech, with the lecturer standing before a lectern on a platform facing a large audience.

The lecture consisted mainly of a narrative supported by a few PowerPoint slides that were presented on a large screen visible to the audience. The recording primarily displayed the speaker and tended to switch briefly to slide view when a new issue was brought up. The lecture dealt with the topic of "cyber illness". It addressed the health risks of digitalization, especially for the development of young people. It was chosen because it was deemed an engaging presentation on a topic that was presumed to interest the participants.

The whole lecture lasted 28 min, 26 s. It was split into sections to create room for the embedded questions. Based on what seemed meaningful event boundaries, this led to four separate video sections: video 1 (7 min, 28 s), video 2 (8 min, 48 s), video 3 (7 min, 2 s), and video 4 (5 min, 8 s). Video 1 introduced the term "smomby," which refers to a smartphone zombie. It explained the speaker's claim that extensive exposure to phones, games and computers can cause serious physical and emotional health problems, including reduced empathy. Video 2 briefly discussed a slide with an overview of brain development over the course of a lifetime (see Fig. 1). The narrative that followed mainly concentrated on early language and sensorimotor development. Video
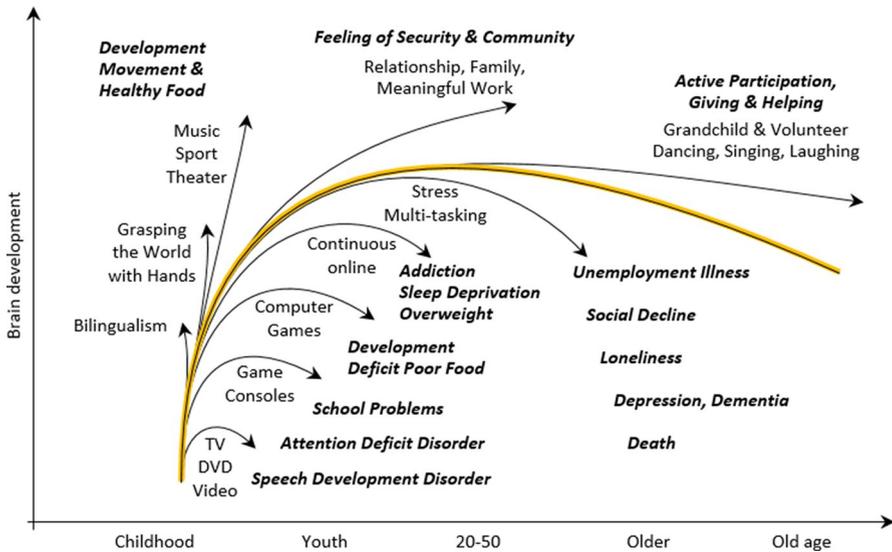
Fig. 1 Slide with the theoretical model in the lecture (translated version)

3 discussed two experiments that showed the negative effect of computer-based compared to tactile learning by young children. Video 4 discussed brain development at the other end of the age scale. The narrative concentrated on dementia, its consequences and antecedents. It ended in the speaker's plea for brain training.

The embedded questions included in the experimental condition were: (Q1) What skill diminishes when people spend a long time viewing a computer screen? (Q2) What are the key dimensions that are responsible for how well our brain develops? (Q3) What do babies do when they see something that does not fit with what they know? (Q4) What are the five technical features that are connected with reduced brain development? These questions all addressed a key aspect of the video segment that they followed. For instance, Q2 asked about an important aspect of the theoretical model featured in video 2. The answer is the three factors presented in bold across the top in Fig. 1.

Videos in the experimental condition ended with an automatically presented embedded question. Videos in the control condition simply ended at the end of each segment. Participants needed to select the next video to move the lecture forward. Both conditions saw the lecture in four videos rather than as a whole, to avoid confounding question-asking and segmentation (see Cheon et al. 2014; Spanjers et al. 2012a, b).

Research instruments

The lecture was presented on a specially created website connected to a *logging instrument* that recorded time-stamped viewer actions for each video. Three engagement measures were gathered: Basic play time, total time and replays. *Basic play*

time was the percentage of unique video seconds set into play mode. A score of 100% for basic play provides tentative evidence that a video has been viewed in full, insofar as it has at least been played through in its entirety. *Total time* was the mean amount of time participants spent on each video. The measure (in s) included pauses. Due to a software glitch, total time could be computed only for the first three videos. Therefore, the comparison between conditions for this engagement measure did not include the time spent on the fourth video. *Replays* were operationalized as actions that follow after an initial viewing of the complete video, in which the user returns to an earlier part of the video and plays a segment of the video again before closing it to move to the next video or the question. Replays are likely to be affected by embedded questions and are a signal of restudying activities. Both the frequency and the duration of replays were measured.

A *paper questionnaire* measured technology acceptance and self-efficacy ratings. Its construction was based on the original questionnaires constructed by respectively Davis (1989) and Vollmeyer and Rheinberg (2006) with modifications to fit the specific context of the study. The questionnaire consisted of a total of 30 statements. There were six distractor items, and six items per construct. *Usefulness* was defined as the degree to which a person generally believes that viewing recorded lectures enhances learning (compare Davis 1989). Examples of usefulness statements are "Recorded lectures like these are useful for studying" and "Students benefit from having recorded lectures available." *Ease of use* was defined as the degree to which a person generally believes that viewing recorded lectures is relatively effortless (compare Davis 1989). For ease of use, statements such as "Recorded lectures require less effort to follow than real lectures" and "Recorded lectures are easy to use" were presented. One item for this construct correlated poorly with the others and was dropped from further analyses. *Satisfaction* was defined as the degree to which a person experiences a positive emotion from viewing a specific recorded lecture (compare Joo et al. 2014; Shin et al. 2011). Satisfaction was measured with statements such as "I enjoyed viewing the video" and "It was a satisfying experience to view the video."

For *self-efficacy,* statements about retention and comprehension of the content of the video-recorded lecture were presented (e.g., "I can write a good summary of the recorded lecture" and "I can remember the content of the recorded lecture quite well"). Responses indicating degree of agreement with each statement were given on a 7-point Likert scale, with scale values that ranged from *completely disagree* (1) to *completely agree* (7). Reliability analyses showed that there were satisfactory to good Cronbach's alpha scores for the four constructs (usefulness = 0.81; ease of use = 0.66; satisfaction = 0.93; self-efficacy = 0.80).

A computer-based *knowledge test* measured retention and comprehension of the lecture. The test was presented on the same website as the videos and consisted of 6 open-ended, brief response items that asked for facts or concepts. Only one test item was the same as an embedded question (i.e., Q2 is the same as T4). The test questions were: (T1) What percentage of learning loss was found in a study where W-lan was installed in the classroom? (T2) What examples illustrate the effect of "smomby" behavior on young people's character? Please also mention how bystanders reacted. (T3) What percentage of the brain is underused when motoric tasks are

viewed on a computer screen as opposed to actual manipulation? (T4) What are the key dimensions that are responsible for how well our brain develops? [This item repeated the second embedded question] (T5) What types of knowledge should be tested to prove that babies need to feel rather than view on a computer screen? (T6) Five technology or technology-related aspects were mentioned for which extensive use, or exposure, could have negative consequences. These aspects were: (1) TV, DVD and video, (2) arcade games, (3) computer games, (4) continuously being online, (5) stress and multitasking. Mention as many of these consequences for each aspect that you can.

Just like the embedded questions, the test items referred to important lecture content from each section of the overall video, and with each of the four parts mentioned in at least one item. For instance, items T4 and T6 concerned the theoretical model. Item T5 asked for the kinds of knowledge tested in an extensively discussed experiment. A codebook provided clearly defined correct and incorrect responses and there were no difficulties in identifying them as such when scoring the responses. Items varied in the number of points that could be obtained. The score for each item was converted to the percentage of possible points obtained, and the overall test score is the mean percentage for all items on the knowledge test.

Procedure

The experiment took place in a small room that seated four participants at a time (all from the same condition). Each participant worked on a laptop with a touchpad and mouse and wore earphones during the experiment. The experimenter told participants that they would view a recorded lecture consisting of four short video segments and that they would be tested on what they understood and remembered. Participants in the experimental condition were also alerted to the presence of questions at the end of each video. They were told that the questions could be used to prepare themselves for the knowledge test. The participants were told to view the videos one after the other in the indicated sequence. They could process each video as they wanted as long as it was open, but they were not allowed to revisit a video they had closed. Note-taking was not allowed. After viewing all videos, participants first completed the questionnaire (on paper) and then took the knowledge test (on the laptop).

Data analysis

Tests revealed that the control and experimental condition had the same distribution for gender (i.e., 5 males and 15 females) and did not differ in age. Assumption testing revealed violations of the normality distribution for the video engagement measures. Therefore, an effect of condition on scores on those measures was assessed with a Mann–Whitney test. ANOVAs could be used for the questionnaire and knowledge test data. Testing was two-tailed with α set at 0.05. For effect sizes, the r-statistic is reported (Field 2013). This statistic tends to be qualified as small, medium, and large for the values $r = 0.10$, $r = 0.30$, and $r = 0.50$, respectively.

Results

What is the effect of condition on video engagement?

Analyses for *basic plays* yielded scores of 100% (or very close) for all videos and participants, indicating that all videos in both conditions were played at least once in full. The experimental group had a significantly higher *total time* score ($Mdn = 26$ min) than the control group ($Mdn = 24$ min 18 s), $U(40) = 52.00$, $z = 4.01$, $p < 0.001$, $r = 0.63$. Also, the experimental group had a significantly higher number ($Mdn = 1.83$) and duration ($Mdn = 10.67$ s) of *replays* than the control group ($Mdn = 0.00$; $Mdn = 0.00$). For number of replays, $U(40) = 58.50$, $z = 4.23$, $p < 0.001$, $r = 0.67$; for duration of replays, $U(40) = 56.50$, $z = 4.38$, $p < 0.001$, $r = 0.69$.

What is the effect of condition on technology acceptance and self-efficacy?

Table 1 shows the mean scores for the technology acceptance constructs and self-efficacy. The scores were uniformly positive and high, lying almost 2 standard deviations above the mid-scale value of 4. There were no differences between conditions, with all F-values < 1.00.

What is the effect of condition on knowledge development?

Table 2 shows the mean scores on the knowledge test. There was an overall effect of embedded questions on learning, $F(1, 39) = 4.40$, $p = 0.043$, $r = 0.32$. However, on closer inspection, this effect was limited to the one item (T4) that asked for information that was tested in an embedded question (Q2), $F(1, 39) = 4.20$, $p = 0.047$, $r = 0.31$. There was a positive but non-significant effect of embedded questions on the items that asked for other information that had not already been tested in the embedded questions, $F(1, 39) = 1.81$, $p = 0.19$.

Exploration of the relationships between engagement (total time and replays-duration) and the knowledge test scores yielded low, non-significant rank correlations overall, as well as within conditions.

Table 1 Means (SDs) for self-efficacy, usefulness, ease of use, and satisfaction per condition

	Self-efficacy M (SD)	Usefulness M (SD)	Ease of use M (SD)	Satisfaction M (SD)
Experimental ($n = 20$)	5.46 (0.63)	5.45 (0.79)	5.47 (0.65)	5.56 (0.69)
Control ($n = 20$)	5.51 (0.85)	5.51 (0.97)	5.45 (0.82)	5.46 (1.21)
Total ($n = 40$)	5.48 (0.74)	5.48 (0.87)	5.46 (0.74)	5.51 (0.97)

Scale values range from 1 to 7, with higher values meaning a more positive rating

Table 2 Means (SDs) for the knowledge test per condition

	Knowledge test		
	Quizzed items M (SD)	Non-quizzed items M (SD)	Overall M (SD)
Experimental ($n=20$)	57.5% (46.7)	49.3% (14.6)	50.8% (15.8)
Control ($n=20$)	30.0% (37.7)	43.7% (11.6)	41.4% (11.9)
Total ($n=40$)	43.8% (44.1)	46.5% (13.4)	46.1% (14.6)

Discussion and conclusion

Video engagement

Participants engaged significantly and substantially longer with the video-recorded lecture that included embedded questions. The presence of quizzing resulted in higher overall scores for total time and replays. The findings are aligned with the results in two empirical studies that also reported that more time is spent on lectures with embedded questions compared to non-quizzed lectures (Kovacs 2016; Vural 2013).

Total time is a general signal of participant interactions with the video. In automated data analyses of video usage, total time is considered one of the most important signals of active processing (Guo et al. 2014). The data on replays complement these data. The finding that these replays were both more frequent and longer in duration with the presence of questions matches our expectations, but even so, merely attests to the effectiveness of questions as a stimulus for video processing.

Logging systems enable researchers to mine a variety of user interactions with the video. In this study, these measures were restricted to general records of video processing (i.e., basic play and total time), and a specific record of video processing that is a probable signal of a remediating action (i.e., replays). Future research might want to use more refined data mining techniques to probe more deeply into the effects of embedded questions on video interaction events. For instance, records could be examined for the number of users who attempt to answer the embedded questions, the correctness of the answers, whether replays occur before or after giving an answer, and the time spent on answering (e.g., Kovacs 2016; Li et al. 2015; Li and Baker 2018). Such information can provide answers to questions such as whether embedded questions affect video navigation, and whether there are interaction event peaks around these questions.

Technology acceptance and self-efficacy

The mean scores for usefulness, ease of use and satisfaction were about 5.5 on a 7-point scale in each condition. Participants believed that video-recorded lectures generally provide a useful resource for learning and are easy to process. In addition,

they felt that the specific lecture yielded a satisfying experience. These findings are in line with a large number of studies that have reported positive student appraisals of video-recorded lectures (e.g., Baepler et al. 2014; Burgoyne and Eaton 2018; Kim et al. 2014).

There was no effect of condition on usability perceptions. The findings thus did not support the positive contribution of quizzing to satisfaction reported by Spanjers et al. (2015). The absence of such a contribution could have been due to the positive overall appraisals of the lecture itself. Voluntary comments from participants after the experiment indicated that they enjoyed the topic and how it was presented. These positive comments may have overshadowed any perceived benefits of quizzing within the lecture.

In both conditions, the self-efficacy score was considerably above the neutral midscale value. This suggests that participants were fairly confident about their knowledge development. They believed that they remembered and understood the video-recorded lecture well. The positive appraisals for self-efficacy hold considerable promise for students' future engagement with video-recorded lectures. That is, current self-efficacy has been found to be a predictor of future persistence and effort expenditure in comparable settings (Bandura 2012; Bandura and Locke 2003).

There was no effect of condition on self-efficacy. This finding thus did not corroborate the outcome of a recent meta-analysis on clickers (Hunsu et al. 2016), nor did it replicate the laboratory study by Tweissi (2016) that found a significant effect of embedded questions on self-efficacy. One possible explanation is that, unlike in the present study, in most clicker studies, as well as in Tweissi's research, feedback was given for the responses to the questions. Another explanation is that the intervention may have been too short to influence the students' self-efficacy. It takes time to hone ones' skills, and increase confidence in grasping the content of a video-recorded lecture; a single, 30 min lecture is unlikely to effect a major change in these facets. Future research might therefore want to investigate whether repeated exposure to video-recorded lectures with embedded questions benefits students' self-efficacy development more from than repeated exposure to such lectures without questions.

Knowledge development

The presence of the embedded questions had a significant, medium-sized effect on what was learned from the lecture. By and large, this concurs with the findings on adjunct questions and on quizzing (e.g., Jing et al. 2016; Smith et al. 2010; Uner and Roediger 2018; Vural 2013). To our knowledge, the present study is one of the few controlled experiments on open-ended questions in video-recorded lectures, and it is the only study in which no feedback was given for responses to these questions (see Cummins et al. 2016; Szpunar et al. 2013, 2014; Thomas et al. 2018). The absence of feedback, in combination with asking open-ended instead of multiple-choice questions, was considered to be a strong stimulus for students to engage in information retrieval, which would thereby enhance their learning.

Dunlosky et al. (2013) have argued that open-ended questions are more likely to trigger elaborate retrieval processes than multiple-choice questions do. That is,

their review of learning techniques indicated that practice tests that require more generative responses (such as recall or short-answer) are more effective than tests that require less generative responses (such as recognition). They also mentioned, however, that this conclusion is tentative and that further work is needed. Karpicke's (2017) more recent review indicated that comparative studies from the last 10 years have yielded mixed outcomes and that finding proof for this claim is more complex than initially thought. Among other things, he pointed out that initial retrieval success can play a mediating role, because it is often higher for multiple-choice questions. In addition, he mentioned that feedback plays a mediating role and that the presence of feedback seems especially relevant for open-ended questions.

The present study found that even without feedback, the presence of open-ended questions enhanced learning. An important reason for the absence of feedback was that we wanted to prevent the risk that students might engage less in constructing their own answers when feedback was present. That is, research on adjunct questions warns that feedback can forestall the students' attempt to retrieve or reconstruct the answer from memory, as students are inclined to depend more on the feedback as a way to obtain the correct answer (Hamaker 1986; Roelle et al. 2017). In addition, a recent meta-analysis on the testing effect provided a similar explanation for the surprising finding that feedback did not moderate learning (Adesope et al. 2017). In view of the substantial evidence in favor of feedback (e.g., Fiorella and Mayer 2018; Shute 2008), Adesope et al. (2017) suggested that more research is needed to reveal when feedback does or does not enhance learning. One feature of feedback that the meta-analysis did not have enough research on to analyze was timing (i.e., immediate versus delayed). To enhance the effectiveness of quizzing, future studies might want to investigate the contribution of delayed feedback, because this design can serve the dual goal of both stimulating and supporting the students' thought processes, using a different time point for each. That is, the absence of immediate feedback for responses to the open-ended questions may keep the students challenged to construct their own answers, while information about the correct response that is available after a delay may help substantiate, enrich or challenge their own constructed answers in a productive way.

The logged records of the students' actions involving video play revealed that the presence of questions led to more extensive playing, and hence potential viewing, of the videos. Unfortunately, these data could not be linked to the students' answers to the embedded questions, because they were not recorded in the present study. This seems like an important issue for further research. If student answers to the embedded questions are known and (delayed) feedback is given, then plausible effects of both of these on additional video play can be evaluated (e.g., correct responses yield little or no extra engagement, incorrect responses stimulate repeated video play). A simpler set-up that such a study could use would be to pre-assess the difficulty level of the questions and then correlate that with video engagement.

Video engagement was not related to test performance. This outcome was unexpected, because empirical studies generally report that more video engagement leads to higher learning outcomes (e.g., Morris et al. 2005; Wei et al. 2015).

The finding stresses the point that video engagement is a proxy for video processing. It is a valuable, unobtrusive record that is a necessary but not sufficient

prerequisite for comprehension and learning. As mentioned earlier, more refined records of video interaction events can provide a more detailed view on the effects of embedded questions on video engagement. Future research might want to complement these records with interview data to gather information on the reasons why users do, or do not, engage with embedded questions (e.g., Shin et al. 2018). Such studies could also considered recording verbal protocols or use other observational methods to obtain insights into how users process embedded questions. Such data could reveal whether the questions prompt users to reflect on the lecture and whether they connect the new information with prior knowledge, among others.

The experimental condition had a significantly higher score for the single repeated question, but did not differ from the control condition on the remaining composite test score with that question removed. This finding indicates that while embedded questions have a moderate effect on learning of questioned content, effects on learning of non-questioned content may be more limited. A similar cautionary note has been voiced for quizzing in applied settings using authentic educational materials (Agarwal et al. 2012; Nguyen and McDaniel 2015; Wooldrige et al. 2014). Future research might therefore want to test this by systematically varying the number and type of previously questioned or non-questioned items. Such research might also contribute to gathering information on the learning strategies involved in video replays. In addition, it might want to measure students' self-regulated learning skills, as this appears to be an ignored moderator in research on quizzing (see Shapiro et al. 2017).

Some limitations of the study have already been mentioned, such as incomplete data on total time, and the absence of information about the answers to the embedded questions. One other limitation has not yet been mentioned, namely, the topic of segmentation. Embedded questions obviously need to be positioned somewhere during a video-recorded lecture. The issue is where such breaks can best be created. By their very nature, embedded questions break down a lecture into parts. This can induce a segmenting effect (see Mayer and Pilegard 2014). To disentangle the two factors (i.e., questions and segmentation), in the present study, both conditions received separate parts of the complete lecture. The video-recorded lecture was split into four segments (videos) and each new segment required a user action to start it playing, as recommended in the multimedia literature (e.g., Biard et al. 2018). In the control condition, each video simply ended when an event was completed; in the experimental condition, there was an open-ended question. Given this parallel construction, it is possible to draw conclusions about the effect of embedded questions. Since embedded question automatically split up a complete lecture into sections, future research on their effects might want to turn to the multimedia literature for a principle-based approach to creating meaningful segments (e.g., Khacharem et al. 2013; Mura et al. 2013; Spanjers et al. 2012b).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval A research proposal describing the study has been submitted to the ethics committee of the University who has given approval.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306.

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,* 437–448. https://doi.org/10.1007/s10648-012-9210-2.

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

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 89–132). New York, NY: Academic Press.

Andre, T. (1981). The role of paraphrased adjunct questions in facilitating learning from prose. *Contemporary Educational Psychology, 6,* 22–27.

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers and Education, 78,* 227–236. https://doi.org/10.1016/j.compedu.2014.06.006.

Baker, P. R. A., Demant, D., & Cathcart, A. (2018). Technology in public health higher education. *Asia-Pacific Journal of Public Health, 30*(7), 655–665. https://doi.org/10.1177/1010539518800337.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: Freeman and Company.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9–44. https://doi.org/10.1177/0149206311410606.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87–99. https://doi.org/10.1037/0021-9010.88.1.87.

Barr, M. L. (2017). Encouraging college student active engagement in learning: Student response methods and anonymity. *Journal of Computer Assisted learning, 33,* 621–632. https://doi.org/10.1111/jcal.12205.

Batchelor, J. (2015). Effects of clicker use on calculus students' mathematics anxiety. *PRIMUS, 25*(5), 453–472. https://doi.org/10.1080/10511970.2015.1027976.

Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior, 89,* 411–417. https://doi.org/10.1016/j.chb.2017.12.002.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education, 151,* 1–36.

Brink, A. G. (2013). The impact of pre- and post-lecture quizzes on performance in intermediate accounting II. *Issues in Accounting Education, 28*(3), 461–485. https://doi.org/10.2308/iace-50445.

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2016). Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance. *IEEE Transactions on Signal Processing, 64*(14), 3677–3692. https://doi.org/10.1109/tsp.2016.2546228.

Buil, I., Catalan, S., & Martinez, E. (2016). Do clickers enhance learning? A control-value theory approach. *Computers and Education, 103,* 170–182. https://doi.org/10.1016/j.compedu.2016.10.009.

Burgoyne, S., & Eaton, J. (2018). The partially flipped classroom: The effects of flipping a module on "Junk Science" in a large methods course. *Teaching of Psychology, 45*(2), 154–157. https://doi.org/10.1177/0098628318762894.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4–5), 514–527. https://doi.org/10.1080/09541440701326097.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education, 6*(1), 9–20.

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. https://doi.org/10.1037/a0024140.

Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied, 24*(1), 34–42. https://doi.org/10.1037/xap0000145.

Cheon, J., Chung, S., Crooks, S. M., Song, J., & Kim, J. (2014). An investigation of the effects of different types of activities during pauses in a segmented instructional animation. *Educational Technology and Society, 17*(2), 296–306.

Christiansen, M. A., Lambert, A. M., Nadelson, L. S., Dupree, K. M., & Kingsford, T. A. (2017). In-class versus at-home quizzes: Which is better? A Flipped learning study in a two-site synchronously broadcast organic chemistry course. *Journal of Chemical Education, 94*(2), 157–163.

Crandall, P. G., Clark, J. A., Shoulders, C. W., & Johnson, D. M. (2019). Do embedded assessments in a dual-level food chemistry course offer measurable learning advantages? *Journal of Food Science Education, 18,* 67–70. https://doi.org/10.1111/1541-4329.12159.

Cummins, S., Beresford, A. R., & Rice, A. (2016). Investigating engagement with in-video quiz questions in a programming course. *IEEE Transactions on Learning Technologies, 9*(1), 57–66. https://doi.org/10.1109/TLT.2015.2444374.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted learning, 20,* 81–94.

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.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior, 89,* 465–470. https://doi.org/10.1016/j.chb.2018.07.015.

Fredriksen, H. (2020). Exploring realistic mathematics education in a flipped classroom context at the tertiary level. *International Journal of Science and Mathematical Education*. https://doi.org/10.1007/s10763-020-10053-1.

Garcia-Rodicio, H. (2015). Questioning as an instructional strategy in multimedia environments: Does having to answer make a difference? *Journal of Educational Computing Research, 52*(3), 365–380. https://doi.org/10.1177/0735633115571931.

Gier, V. S., & Kreiner, D. S. (2009). Incorporating active learning with PowerPoint-based lectures using content-based questions. *Teaching of Psychology, 36,* 134–139. https://doi.org/10.1080/00986280902739792.

Guo, P. J., Kim, J., & Rubin, R. (2014). *How video production affects student engagement: An empirical study of MOOC videos.* Paper presented at the L@S '14, Atlanta, GA.

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*. https://doi.org/10.1007/s10956-020-09847-3.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*(2), 212–242. https://doi.org/10.2307/1170376.

Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55*(1), 47–85. https://doi.org/10.2307/1170407.

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers and Education, 94,* 102–119. https://doi.org/10.1016/j.compedu.2015.11.013.

inCITI Singen. (2015). *Cyberkrank—Wie das digitale Leben unsere Gesundheit ruiniert* [Cyber illness—How our digital life ruins our health]. [Video] YouTube. https://www.youtube.com/watch?v=9SrVF9vXHyU&t=1847s.

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.

Jolley, D. F., Wilson, S. R., Kelso, C., O'Brien, G., & Mason, C. E. (2016). Analytical thinking, analytical action: Using prelab video demonstrations and e-quizzes to improve undergraduate preparedness for analytical chemistry practical classes. *Journal of Chemical Education, 93,* 1855–1862. https://doi.org/10.1021/acs.jchemed.6b00266.

Joo, Y. J., Lee, H. W., & Ham, Y. (2014). Integrating user interface and personal innovativeness into the TAM for mobile learning in Cyber University. *Journal of Computing in Higher Education, 26*(2), 143–158. https://doi.org/10.1007/s12528-014-9081-2.

Karabulut-Igu, A., Cherrez, N. J., & Jahren, C. T. (2018). A systematic review of research on the flipped classroom method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411. https://doi.org/10.1111/bjet.12548.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 487–514). Amsterdam: Academic Press.

Ketsman, O., Daher, T., & Santana, J. A. C. (2018). An investigation of effects of instructional videos in an undergraduate physics course. *E-Learning and Digital Media, 15*(6), 267–289. https://doi.org/10.1177/2042753018805594.

Khacharem, A., Spanjers, I. A. E., Zoudji, B., Kalyuga, S., & Ripoll, H. (2013). Using segmentation to support learning from animated soccer scenes: An effect of prior knowledge. *Psychology of Sports and Exercise, 14,* 154–160.

Khan, A., Schoenborn, P., & Sharma, S. (2019). The use of clickers in instrumentation and control engineering education: A case study. *European Journal of Engineering Education, 44*(1–2), 271–282. https://doi.org/10.1080/03043797.2017.1405240.

Khanna, M. M. (2015). Ungraded pop quizzes: Test-enhanced learning without all the anxiety. *Teaching of Psychology, 42*(2), 174–178. https://doi.org/10.1177/0098628315573144.

Kim, M. K., Kim, S. O., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *Internet and Higher Education, 22,* 37–50. https://doi.org/10.1016/j.iheduc.2014.04.003.

Kovacs, G. (2016). *Effects of in-video quizzes on MOOC lecture viewing.* Paper presented at the third ACM conference on learning @ scale, Edinburgh, Scotland, UK.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education, 31*(1), 30–43.

Lawson, T. J., Bodle, J. H., Houlette, M. A., & Haubner, R. R. (2006). Guiding questions enhance student learning from educational videos. *Teaching of Psychology, 33*(1), 31–33.

Leisner, D., Zahn, C., Ruf, A., & Cattaneo, A. (2020). *Different ways of interacting with videos during learning in secondary physics lessons.* Paper presented at the 22nd International Conference on Human–Computer Interaction, HCII 2020, Copenhagen, Denmark.

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers and Education, 127,* 41–65. https://doi.org/10.1016/j.compedu.2018.08.005.

Li, N., Kidziński, L., Jermann, P., & Dillenbourg, P. (2015). *MOOC video interaction patterns: What do they tell us?* Paper presented at the 10th European Conference on Technology Enhanced Learning (EC-TEL), Toledo, Spain.

Lin, Y.-T., & Chen, C.-M. (2019). Improving effectiveness of learners' review of video lectures by using an attention-based video lecture review mechanism based on brainwave signals. *Interactive Learning Environments, 27*(1), 86–102. https://doi.org/10.1080/10494820.2018.1451899.

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). New York, NY: Cambridge University Press.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34,* 51–57. https://doi.org/10.1016/j.cedpsych.2008.04.002.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103*(2), 399–414.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4–5), 494–513. https://doi.org/10.1080/09541440701326154.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27,* 360–372. https://doi.org/10.1002/acp.2914.

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition, 1,* 18–26. https://doi.org/10.1016/j.jarmac.2011.10.001.

McDermott, K. B., Agarwal, P. K., D'Antonio, L. D., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied, 20*(1), 3–21. https://doi.org/10.1037/xap0000004.

Morling, B., McAuliffe, M., Cohen, L., & DiLorenzo, T. (2008). Efficacy of personal response systems ("clickers") in large, introductory psychology classes. *Teaching of Psychology, 35,* 45–50. https://doi.org/10.1080/00986280701818516.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *Internet and Higher Education, 8,* 221–231. https://doi.org/10.1016/j.iheduc.2005.06.009.

Mura, K., Petersen, N., Huff, M., & Ghose, T. (2013). IBES: A tool for creating instructions based on event segmentation. *Frontiers in Psychology, 4,* 1–14. https://doi.org/10.3389/fpsyg.2013.00994.

Nguyen, K., & McDaniel, M. A. (2015). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology, 42*(1), 87–92. https://doi.org/10.1177/0098628314562685.

O'Callaghan, F. V., Neumann, D. L., Jones, L., & Creed, P. A. (2017). The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues. *Educational Information Technology, 22,* 399–415. https://doi.org/10.1007/s10639-015-9451-z.

Ozgungor, S., & Guthrie, J. T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. *Journal of Educational Psychology, 96*(3), 437–443. https://doi.org/10.1037/0022-0663.96.3.437.

Pan, S. C., Cooke, J., Little, J. L., McDaniel, M. A., Foster, E. R., Connor, L. T., et al. (2019). Online and clicker quizzing on jargon terms enhances definition-focused but not conceptually focused biology exam performance. *CBE Life Sciences Education, 18,* 1–2. https://doi.org/10.1187/cbe.18-12-0248.

Papadopoulos, P. M., Natsis, A., Obwegeser, N., & Weinberger, A. (2018). Enriching feedback in audience response systems: Analysis and implications of objective and subjective metrics on students' performance and attitudes. *Journal of Computer Assisted learning, 35*(2), 305–316. https://doi.org/10.1111/jcal.12332.

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review, 24,* 419–435. https://doi.org/10.1007/s10648-012-9203-1.

Rice, P., Beeson, P., & Blackmore-Wright, J. (2019). Evaluating the impact of a quiz question within an educational video. *Tech Trends, 63*(5), 522–532. https://doi.org/10.1007/s11528-019-00374-6.

Roelle, J., Rahimkhani-Sagvand, N., & Berthold, K. (2017). Detrimental effects of immediate explanation feedback. *European Journal of Psychology of Education, 32,* 367–384. https://doi.org/10.1007/s10212-016-0317-6.

Rothkopf, E. Z. (1970). The concept of mathemagenic activities. *Review of Educational Research, 40*(3), 325–336.

Shapiro, A. M. (2009). An empirical study of personal response technology for improving attendance and learning in a large class. *Journal of the Scholarship of Teaching and Learning, 9*(1), 13–26.

Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology, 26,* 635–643. https://doi.org/10.1002/acp.2843.

Shapiro, A. M., & Gordon, L. T. (2013). Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning. *Journal of Teaching and Learning with Technology, 2*(1), 15–30.

Shapiro, A. M., Sims-Knight, J., O'Rielly, G. V., Capaldo, P., Pedlow, T., Gordon, L., et al. (2017). Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning. *Computers and Education, 11,* 44–59. https://doi.org/10.1016/j.compedu.2017.03.017.

Shin, H., Ko, E.-Y., Williams, J. J., & Kim, J. (2018). *Understanding the effects of in-video prompting on learners and instructors*. Paper presented at the Conference on Human Factors in Computing Systems (CHI), Montreal, Canada.

Shin, D. H., Shin, Y. J., Choo, H., & Beom, K. (2011). Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices. *Computers in Human Behavior, 27*(6), 2207–2214. https://doi.org/10.1016/j.chb.2011.06.017.

Shinaberger, L. (2017). Components of a flipped classroom influencing student success in an undergraduate business statistics course. *Journal of Statistics Education, 25*(3), 122–130. https://doi.org/10.1080/10691898.2017.1381056.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795.

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching, 47*(4), 363–379. https://doi.org/10.1002/tea.20378.

Spanjers, I. A. E., Könings, K. D., Leppink, J., Verstegen, D. M. L., de Jong, N., Czabanowska, K., et al. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review, 15,* 59–74. https://doi.org/10.1016/j.edurev.2015.05.001.

Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2012a). Segmentation of worked examples: Effects on cognitive load and learning. *Applied Cognitive Psychology, 26,* 352–358. https://doi.org/10.1002/acp.1832.

Spanjers, I. A. E., van Gog, T., Wouters, P., & van Merriënboer, J. J. G. (2012b). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers and Education, 59,* 274–280. https://doi.org/10.1016/j.compedu.2011.12.024.

Strelan, P., Osborn, A., & Palmer, E. (2020). Student satisfaction with courses and instructors in a flipped classroom: A meta-analysis. *Journal of Computer Assisted learning, 36,* 295–314. https://doi.org/10.1111/jcal.12421.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications for online education. *Journal of Applied Research in Memory and Cognition, 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. *PNAS, 110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110.

Thomas, R. C., Weywadt, C. R., Anderson, J. L., Martinez-Papponi, B., & McDaniel, M. A. (2018). Testing encourages transfer between factual and application questions in an online learning environment. *Journal of Applied Research in Memory and Cognition, 7*(2), 252–260. https://doi.org/10.1016/j.jarmac.2018.03.007.

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018a). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning, 13,* 1–14. https://doi.org/10.1007/s11409-017-9175-0.

Toftness, A. R., Carpenter, S. K., Lauber, S., & Mickes, L. (2018b). The limited effects of prequestions on learning from authentic lecture videos. *Journal of Applied Research in Memory and Cognition, 7,* 370–378.

Trees, A. R., & Jackson, M. H. (2007). The learning environment in clicker classrooms: student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology, 32*(1), 21–40. https://doi.org/10.1080/17439880601141179.

Turan, Z., & Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: A systematic review. *Computer Assisted Language Learning, 33*(5–6), 590–606. https://doi.org/10.1080/09588221.2019.1584117.

Tweissi, A. (2016). *The effects of embedded questions strategy in video among graduate students at a middle eastern university.* (Doctoral dissertation), Ohio University, Athens, OH.

Uner, O., & Roediger, H. L. (2018). The effect of question placement on learning from textbook chapters. *Journal of Applied Research in Memory and Cognition, 7,* 116–122.

van Daele, T., Frijns, C., & Lievens, J. (2017). How do students and lecturers experience the interactive use of handheld technology in large enrolment courses? *British Journal of Educational Technology, 48*(6), 1318–1329. https://doi.org/10.1111/bjet.12500.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review, 18*(3), 239–253. https://doi.org/10.1007/s10648-006-9017-0.

Vural, Ö. F. (2013). The impact of a question-embedded video-based learning tool on e-learning. *Educational Sciences: Theory and Practice, 13*(2), 1315–1323.

Wang, Y. H. (2020). Design-based research on integrating learning technology tools into higher education classes to achieve active learning. *Computers and Education.* https://doi.org/10.1016/j.compedu.2020.103935.

Wei, H. C., Peng, H., & Chou, C. (2015). Can more interactivity improve learning achievement in an online course? Effects of college students' perception and actual use of a course-management system on their learning achievement. *Computers and Education, 83,* 10–21. https://doi.org/10.1016/j.compedu.2014.12.013.

Wooldrige, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory and Cognition, 3,* 214–221. https://doi.org/10.1016/j.jarmac.2014.07.001.

Zhang, D., Zhou, L., Briggs, R. O., & Numaker, J. F., Jr. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management, 43,* 15–27. https://doi.org/10.1016/j.im.2005.01.004.

Zhu, E. (2008). *Teaching with clickers* (p. 22). Occasional Papers: Center for Research on Learning and Teaching.

Hans van der Meij is senior researcher and lecturer in instructional technology at the University of Twente (the Netherlands). His specialization is instructional design (e.g., videos).

Linn Böckmann studied psychology at the University of Twente and now works as a learning therapist.