



# Pop-up Questions Within Educational Videos: Effects on Students' Learning

Marjolein E. Haagsman<sup>1</sup> · Karin Scager<sup>2</sup> · Johannes Boonstra<sup>1</sup> · Margot C. Koster<sup>1</sup>

© The Author(s) 2020

## Abstract

Educational videos are increasingly used to let students prepare lesson material at home prior to in-class activities in flipped classrooms. The main challenge of this teaching strategy is to stimulate students to watch these videos attentively before going to class. This paper describes the use of questions that pop-up within relatively long educational videos of 16 min on average and designed to enhance students' engagement and understanding when preparing for in-class activities. The effects of such pop-up questions on students' learning performance were studied within a flipped course in molecular biology. Students had access to videos with or without a variable set of pop-up questions. The experimental group with pop-up questions showed significantly higher test results compared to the group without pop-up questions. Interestingly, students that answered pop-up questions on certain concepts did not score better on items testing these specific concepts than the control group. These results suggest that merely the presence of pop-up questions enhances students' learning. Additional data from interviews, surveys, and learning analytics suggest that pop-up questions influence viewing behavior, likely by promoting engagement. It is concluded that pop-up questions stimulate learning when studying videos outside class through an indirect testing effect.

**Keywords** Interactive video · Pop-up questions · Flipped classroom · Testing effect · Molecular biology · Undergraduate

## Introduction

Educational videos are regularly used to study information at home in flipped classroom education (Bishop and Verleger 2013). The main idea of this flipped classroom model is that traditional class activities are shifted or “flipped” to activities outside class and vice versa. Thus, students study the lesson content outside class, which is often done with the aid of educational videos. Afterwards, students use the knowledge on a higher cognitive level during in-class activities. The main aim of this setup is that teachers are present when students apply the information and probably when most help is needed.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10956-020-09847-3>) contains supplementary material, which is available to authorized users.

✉ Marjolein E. Haagsman  
M.E.Haagsman@uu.nl

<sup>1</sup> Institute of Education, Department of Biology, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

<sup>2</sup> Department of Education, Faculty of Social and Behavioral Sciences, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

Accordingly, flipped lessons have shown to improve students' test performance within Science, Technology, Engineering and Mathematics (STEM) higher education (Baepler et al. 2014; Gross et al. 2015; Lax et al. 2016; Barral et al. 2018).

Although these results of flipped lessons on learning performance are promising, some challenges remain (Lo and Hew 2017). Herreid and Schiller (2013) described two major challenges for flipped classroom education experienced by STEM teachers. The first, rather practical challenge is that teachers find it hard to obtain or design proper videos suitable for studying the lesson content at home. The second challenge is that some students are not prepared well enough for the in-class activities. This last challenge is a fundamental issue of the flipped classroom model since student preparation is a prerequisite for in-depth in-class activities. The current paper aims to investigate whether students' learning outside the classroom can be improved through video design by using questions that pop-up within educational videos.

Some suggestions for effective video design for learning have been made from the perspective of cognitive theory of multimedia learning (Mayer 2002). This theory suggests that media in learning should prompt cognitive processing of the relevant information without overloading the processing capacity of students. This cognitive process is mainly thought to

be promoted by two conditions in video design: segmentation and signaling (Ibrahim 2011). In this model, segmentation is defined as the division of videos into smaller segments while signaling encompasses visual and aural signs that increase students' focus on the most relevant information. As such, students that watch videos with clear signals and small fragments are expected to have less attention to irrelevant information and will better remember the relevant information (Ibrahim 2011).

One highly promising tool to increase the effectivity of educational videos is the introduction of questions that pop-up within the video (Szpunar et al. 2013, 2014; Cummins et al. 2016; Lavigne and Risko 2018). These questions can either be in the form of single interspaced pop-up questions or as so-called interpolated tests with multiple questions. In either case, questions are included at certain intervals within an educational video and are expected to promote active engagement and thereby learning (Kumar 2010; Brame 2016). When used at regular intervals during the video, these pop-up questions hold both segmenting and signaling functions as advised by Ibrahim (2011). The high interest in pop-up questions as a learning tool is reflected in the number of companies that provide tools to enrich videos with integrated questions (HapYak 2020; Hihaho 2020; H5P 2020; Panopto 2020; PlayPosit 2020; Scalable Learning 2020).

One of the early studies on videos with questions revealed that psychology students achieved higher performance on video-related test questions when watching the video with guiding questions on a separate sheet of paper (Lawson et al. 2006). Furthermore, several studies show that interpolated tests within or between videos do improve students' final test performance (Szpunar et al. 2013; Vural 2013; Lavigne and Risko 2018). The interpolated tests even appeared to be more effective on test performance than extra study time (Szpunar et al. 2013). In contrast, in a study of Wieling and Hofman (2010), interpolated tests did not affect students' final test performance within a course on European Law.

The positive effect of interpolated tests on final test performance observed in some studies could relate to the retrieval or testing effect, which is the finding that taking or practicing tests in general improves retention of information (Glover 1989). This testing effect has been supported by many studies and can be explained by both direct and indirect effects (e.g., Karpicke and Blunt 2011; Pastötter and Bäuml 2014; Roediger and Butler 2011; Szpunar et al. 2008). The direct effect of testing occurs when testing enhances retention on a specific tested topic (Jacoby et al. 2010). One explanation for the direct testing effect is that students need to retrieve and process specific information when doing tests (Roediger and Karpicke 2006). A second possible mechanism for the direct testing effect of pop-up questions in particular is that these questions operate as a signaling tool by recapitulating and testing the most relevant video content.

Besides enhancing retention on the specific tested topic, testing has also shown to enhance retention on subsequent non-tested lesson material (Chan et al. 2006; Szpunar et al. 2008). This indirect testing effect implies that also factors other than re-exposure and retrieval contribute to improved learning performance from tests. Recently suggested mechanisms for indirect testing effects of pop-up questions are an increase in note-taking (Lawson et al. 2006; Szpunar et al. 2013) and spending more time on the online learning material (Vural 2013). Moreover, students have reported to be more focused after each video fragment when they were tested during the videos (Szpunar et al. 2013), suggesting that the questions function as a segmentation tool. Summarizing, previous studies on videos with integrated questions suggest that they might promote learning both directly and indirectly by helping students to focus on the tested and most relevant information, process the tested information more elaborately, retain attention and stay actively involved. The number of studies on pop-up questions is however rather scarce and results are inconclusive.

More insight into the effect of pop-up questions on learning may help teachers to design effective videos. Such insights can be essential for flipped classroom education since the success of this model depends on the preparation by the students. In this study, educational videos on molecular biology are used of about 16 min on average. These rather long educational videos are segmented with questions that pop-up about once per 5 or 6 min. The aim of this paper is to examine whether pop-up questions enhance students' learning outside class within a flipped course in molecular biology. This study specifically aims to address the following three questions:

1. Do students experience that pop-up questions help them in learning the video content?
2. Does the content of pop-up questions result in a direct testing effect?
3. Does the presence of pop-up questions result in an indirect testing effect?

Based on the results from these studies we performed an additional explorative study to address the question:

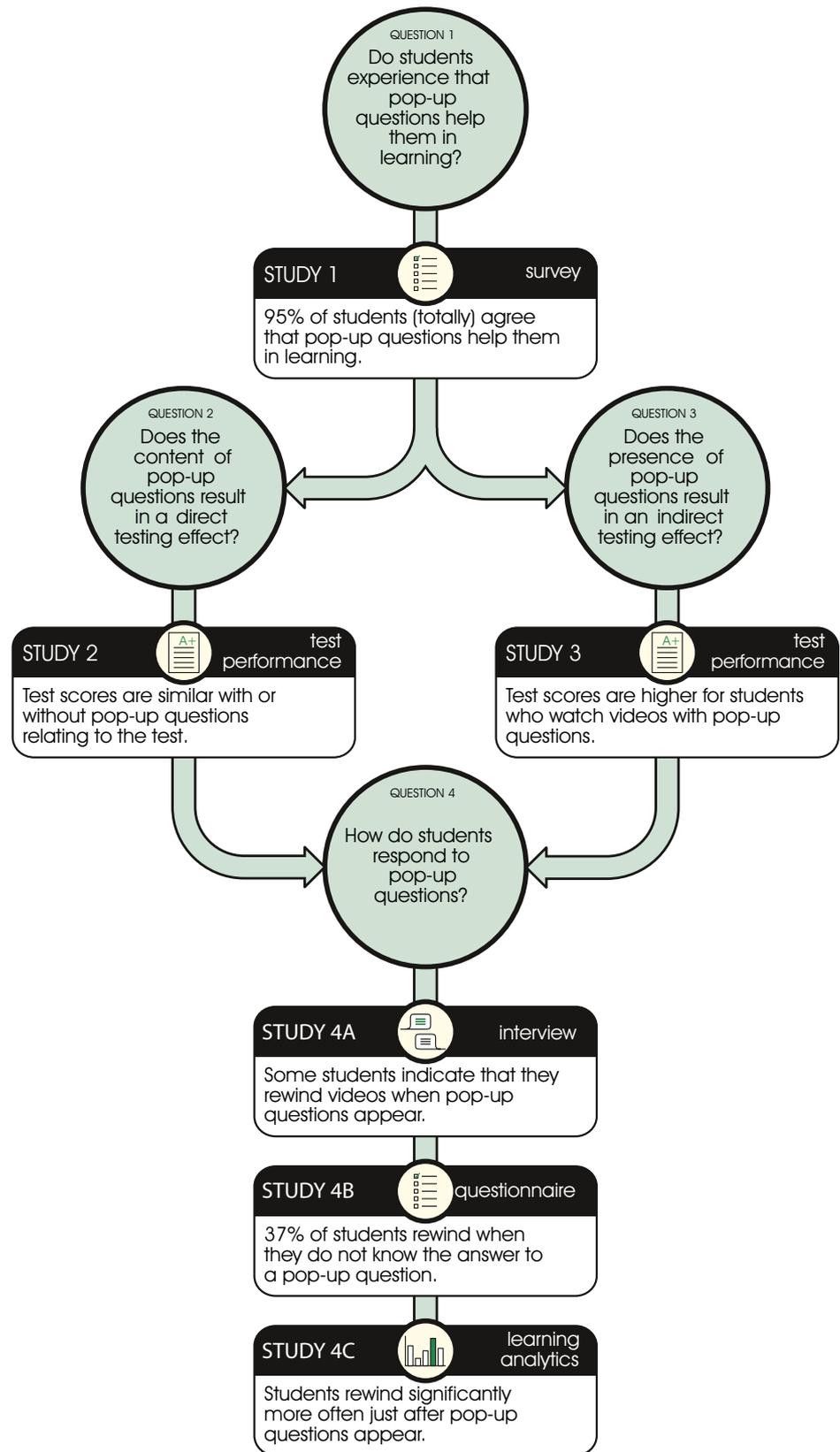
4. How do students use pop-up questions?

The studies were performed in an authentic setting, meaning that the students watched the videos at home, while data were obtained from tests, surveys, interviews, and learning analytics.

## Methods

For this study a multimethod evaluation design was used. Qualitative and quantitative methods were employed

**Fig. 1** Schematic representation of the first three primary studies and final fourth explorative study performed for this research. The connections between the research questions show how the result of one study is used to inform the next



sequentially, using the results of one method to design the next (Fig. 1). First, students’ perception of the effect of interactive

videos on their learning was measured using an evaluation survey in 2015 (Table 1). Second, in 2016, we studied direct

**Table 1** Descriptive statistics for studies on effects of pop-up questions on students' learning

Study	Research question	Instrument	Number of participants, <i>N</i> (%)	Year of data collection
Study 1	Do students experience that pop-up questions help them in learning the video content?	Survey	168 (69%)	2015
Study 2	Does the content of pop-up questions result in a direct testing effect?	Test performance	253 (94%)	2016
Study 3	Does the presence of pop-up questions result in an indirect testing effect?	Test performance	170 (57%)	2017
Study 4A	How do students use pop-up questions?	Interview	14 (8%)	2017
Study 4B	What do students do when pop-up questions appear?	Questionnaire	118 (69%)	2017
Study 4C	Do students rewind the video more often after pop-up questions appear?	Learning analytics	244 (82%)	2017

testing effects of specific pop-up questions on the understanding of corresponding concepts. As a third step in 2017, we studied indirect testing effects from overall test performances.

Based on results from these studies, we performed an extra study using focus group interviews to explore how students make use of pop-up questions. This outcome similarly resulted in a next study to explore students' viewing behavior with the aid of questionnaires. The result of this examination ultimately led to a final study to measure effects of pop-up questions on students' viewing behavior from learning analytics data.

## Participants

The participants in this study were freshman students on the Molecular Biology course (Department of Biology, Utrecht University). The first study on students' perception included 168 participants (69% response rate). The second study on the direct testing effect included 253 participants (94% response rate). The third study on the indirect testing effect was conducted with 170 (57% response rate) participants. The resulting extra studies on viewing behavior from interviews, questionnaires, and learning analytics included respectively 14 (8% response rate), 118 (69% response rate), and 244 participants (82% response rate). Note that the total number of students differs per experiment as the experiments were conducted over a period of 3 years. The participants within the

comparative studies were randomly divided among experimental groups. Descriptive statistics on these groups can be found in Table 2.

## Course Design

The study was performed within the freshman course Molecular Biology, taught at Utrecht University in the Netherlands. The course is given in Dutch and is compulsory for all students participating in the undergraduate program of biology. The course content was based on the Text Book *Biology, A Global Approach*; Chapters 2–13 and Chapters 16–20 (10th and 11th International Edition) (Campbell et al. 2015, 2017). Research on video use was only performed within the first 5 weeks of the course. During this time, students were provided with four to eight videos per week. Students could view the videos voluntarily at home, at their own pace and in their own time. Additionally, understanding of the video content was tested weekly in online tests and then applied in group assignments. All tests, assignments and answers to pop-up questions were discussed weekly with the teacher during obligatory in-class activities in groups of approximately 40 students.

## Video Design

The educational videos were recorded by the teacher of the course. The videos were recorded as screencasts of slides with audio and lasted, on average, 16 min. The topics discussed within the videos were atoms and molecules, chemistry of water, carbon chemistry, biological macromolecules and lipids, energy, cell structure and function, cell membranes, cell signaling, cell cycle, cell respiration, and photosynthesis. The learning goals of the videos are reported in Online Resource 1. The videos were linked to the online video platform ScalableLearning (Scalable Learning) to include pop-up questions within pre-made videos.

## Pop-up Questions Design

Educational videos were enriched with pop-up questions reviewing the previously explained concepts. In 2015, the questions within the videos popped up once per 8 min, on average. In the consecutive year, the main teacher added extra questions to the video as students reported that they would like to have more of them. Extra questions were added up to one question per 5 or 6 min, on average, depending on the experimental group. The questions were designed at the conceptual knowledge level of Bloom's taxonomy (Bloom et al. 1956). The videos paused when pop-up questions appeared within the video and automated feedback was provided after answering the question. Students that viewed the video clip for the first time could only continue the video after the pop-up

**Table 2** Descriptive statistics for comparative studies on effects of pop-up questions on students' learning

Year	Comparative study	Instrument	Group	Response rate, <i>N</i> (%)	%Male	%Female	Average second exam grade (SD) <sup>a</sup>
2016	Study 2	Direct testing effect	Group A	149 (95%)	42%	58%	5.6 (1.50)
			Group B	104 (94%)	38%	62%	5.9 (1.44)
2017	Study 3	Indirect testing effect	Control group	83 (56%)	33%	67%	6.2 (1.52)
			Experimental group	87 (59%)	39%	61%	6.5 (1.46)
	Study 4C	Rewinding behavior	Control group	121 (82%)	47% <sup>b</sup>	53% <sup>b</sup>	6.0 <sup>b</sup> (1.62)
			Experimental group	123 (83%)	43% <sup>b</sup>	57% <sup>b</sup>	6.3 <sup>b</sup> (1.48)

<sup>a</sup> All comparative studies were performed within the first part of the course. The presented grade is the average exam grade for the second part of the course. Students receive exam grades within a range of 1.0 (lowest) to 10.0 (highest)

<sup>b</sup> The demographic information and exam grades for study 4C are given for the total number of students instead of the respondents as data on rewinding behavior are anonymous

question was answered correctly. However, students that rewatched the video clip could continue watching the video at any time by pressing the play button. The number of attempts and correct answers is provided to the teacher for the group as a whole but not per specific student. During the video clip, students also had the opportunity to use additional interactive tools. These tools included making digital notes, asking questions to the teacher and/or fellow students and pressing the “I am confused” button to label video fragments they did not understand. Students could also rewind, fast-forward, pause and change the speed of the video.

### Test Design

Online tests were designed to practice the concepts explained in the videos. The tests were, similarly to the pop-up questions, designed at the comprehension level of Bloom's taxonomy (Bloom et al. 1956). Furthermore, these tests were also used to measure students' learning performance for study 2 and 3 discussed further on. Students were asked to do eight tests of approximately 20 questions each. The tests were performed digitally at home, and the deadline for these tests was 1 day before the corresponding in-class activities. The average score of the eight tests accounted for 5% of the final course grade.

### Study 1—Exploring Students' Experience on the Effect of Pop-up Questions on Their Learning

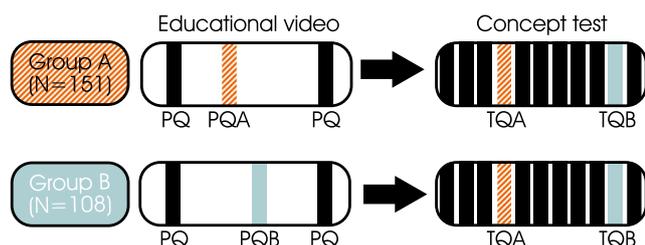
In 2015, videos within the Molecular Biology course were embedded in ScalableLearning for the first time. Students' general perception of (interactive) videos was explored using a survey at the end of the course. Students responded to statements on a 5-point Likert scale, ranging from strongly disagree (1) to strongly agree (5). The survey contained 14 questions on how students used these interactive tools within the video platform and whether these tools affected their learning. Only the following three statements concerning pop-up

questions and the learning effect of educational videos are considered within this paper: *The video clips helped me in learning; Answering the pop-up questions helped me in learning; and I would like to have fewer questions within a video clip.* A translated version of the complete evaluation survey is provided in Online Resource 2.

### Study 2—Measuring the Direct Testing Effect

Students were randomly divided into two groups (A and B) at the start of the course. Each of these groups used a different course environment for watching the educational videos. The videos within the course environments were the same, but 14 extra pop-up questions were inserted for alternating groups (Fig. 2).

The pop-up questions were based on the course learning goals and developed on the level of comprehension. Corresponding test questions were designed for each of those questions and were incorporated into the tests covering the entire study content. The test questions were not identical to the pop-up questions but tested the same concept at the same comprehension level. For example, one pop-up question was: “Which of the following amino acids does not contain asymmetric carbon atoms?” Whereas, the corresponding test question contained a structural formula with the question: “Which of the carbon atoms within the structural formula below is asymmetric?” A translated version of the test questions with the corresponding learning goals is provided in Online Resource 3. Each individual score per specific test question was obtained for comparison. Only test scores were analyzed for students that attempted to answer the corresponding pop-up question. Students who did the digital test before fully watching the corresponding videos were excluded from analysis in order to obtain a solid measurement of the effect of video preparation on test performance. The remaining students did the tests with a median time interval of 1 day 6 h and 58 min after watching the videos.

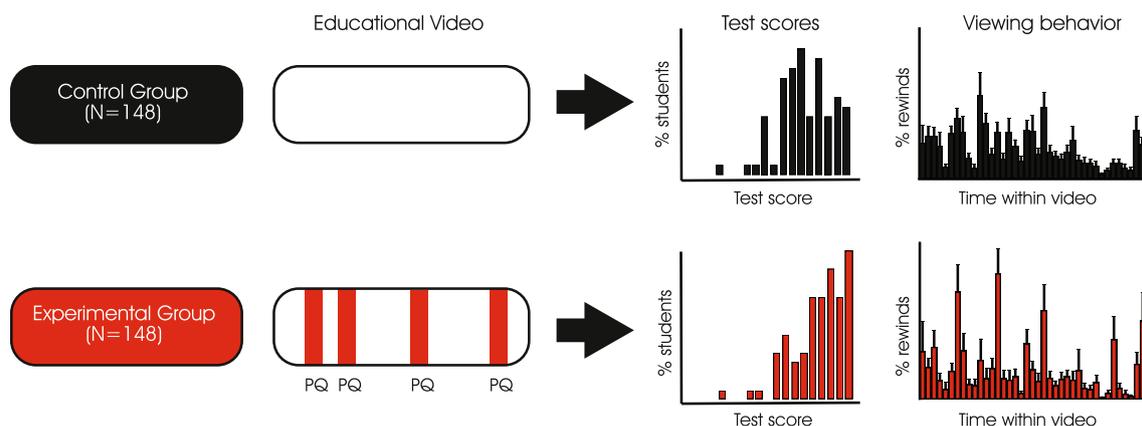


**Fig. 2** Schematic design for measuring understanding of concepts (study 2). In this experiment, students were divided into group A and B. Both groups viewed the same videos with different pop-up questions (PQ). The corresponding test questions (TQ) in the test are marked similarly

### Study 3—Measuring the Indirect Testing Effect

In the subsequent year, students were randomly divided between an experimental and a control group. Both groups watched two educational videos on cell signaling, with durations of, respectively, 20:10 min and 19:33 min. Four pop-up questions were designed for both educational videos. However, in this experiment, only one experimental group received these pop-up questions whereas the other control group received no pop-up questions at all (Fig. 3).

Students' general conceptual understanding of these two videos was tested with a corresponding test on cell signaling. The scores of this test were compared between the control and experimental group and corrected for other test scores obtained prior to the experiment. A translated version of the pop-up questions, test questions, and corresponding learning goals are provided in Online Resource 4. Again, students were excluded from the analysis when they did the digital test before fully watching the corresponding videos.



**Fig. 3** Schematic design for measuring students' overall test performance (study 3) and viewing behavior (study 4C). In this experiment, students were divided into an experimental and a control group. Both groups viewed the same videos on cell signaling. The experimental group received pop-up questions (PQ) within the videos, whereas the control

### Study 4A—Exploring Students' Use of Pop-up Questions via Focus Groups

After the Molecular Biology course in 2016, two groups of six and eight students participated in a semi-structured focus group interview on their use of pop-up questions. Students were asked to describe their actions when questions appeared within the video. A translated version of the guiding questions of the focus group interview is provided in Online Resource 5.

### Study 4B—Exploring Students' Use of Pop-up Questions via Questionnaires

The results of the focus group interviews were used to design a questionnaire on the use of videos and pop-up questions. The questionnaire contained closed questions on students' use of videos, and the multiple-choice answers to these questions were derived from student discussions during the focus group interviews. Only one question concerning students' behavior when not knowing the answer to a pop-up question is used in this paper ( $N = 118$ ). Other questions concerning students' general use of video were considered to be irrelevant for the current paper. A translated version of the questionnaire is provided in Online Resource 6.

### Study 4C—Measuring Students' Rewinding Behavior

Viewing behavior was analyzed for the same video clips on cell signaling (Fig. 3). The specific data used for this study were the number of rewinds per student. Only rewinds of more than 1 s were used for analysis. The percentage of rewinds per student was determined for every timeframe of 30 s

group did not receive any pop-up questions. Students' test performances on the overall video content were compared between the experimental and control group. The groups were also compared on the number of rewinds and fast-forwards per student

in the video. The total use of rewind and fast-forward buttons within the video was also determined for the experimental and control group and compared with a control clip. This control clip was a video clip of 18:07 min on cell structure and function which contained five pop-up questions that were identical for both the control and experimental groups. The raw data on rewinds and fast-forwards were provided personally by the development team of the video platform ScalableLearning.

### Statistical Analysis

For the first step of the analyses, we performed descriptive statistics. The answers to the test questions within study 2 were scored as either correct or incorrect, and Pearson's chi-square analysis was performed to compare these results for groups with or without corresponding pop-up questions. For study 3, a one-way analysis of covariance (ANCOVA) was conducted to compare test scores of the experimental and control group. The test scores were controlled for the second exam grade of the course. An independent *t* test was used for study 4C to compare the mean percentages of rewinds within the 30 s after pop-up questions between the control group and experimental group. The average number of rewinds and fast-forwards throughout the entire videos was not normally distributed and compared with a Mann-Whitney test. Individual rewinds and fast-forwards greater than three times the interquartile range of each experimental group were considered as outliers and removed from this analysis. All statistical analyses were performed using IBM SPSS Statistics Version 24.

## Results

### Student Perception on the Effect of Interactive Video on Their Learning

The present study started with a general student evaluation of the interactive video platform. A few questions within this survey examined whether students believed that educational videos and pop-up questions helped them in learning (Table 3).

Table 3 shows that 97% (totally) agreed that video clips in general helped them in studying the learning content. In addition, 91% of the students (totally) agreed that pop-up questions, specifically, helped them in studying. This positive attitude towards pop-up questions was confirmed by the finding that 79% of the students (totally) disagreed with decreasing the number of questions within the video.

### Direct Testing Effect

Tests were performed prior to the in-class activities to investigate whether a pop-up question on a specific

concept helped students to understand that specific concept. An experimental setup was designed in which two groups watched the same video clips with different pop-up questions on different concepts. Afterwards, both groups did a test on the video content. The test scores on the individual items are shown in Table 4. Surprisingly, the percentage of correctly answered test questions was not significantly different between students that did (72%) or did not (69%) receive corresponding pop-up questions ( $\chi^2(1, N = 2901) = 2.52, p = 0.11$ ). Students with corresponding pop-up questions only performed significantly better on one question (item 5), which was the only question that was nearly identical to the pop-up question ( $\chi^2(1, N = 181) = 15.10, p < 0.001$ ).

### Indirect Testing Effect

In the previous experiment, group A and B were both required to answer pop-up questions although on different concepts. A follow-up experiment was performed to examine whether merely the presence of pop-up questions might affect student performance on the full video content. In this experiment, students watched a video on cell signaling either with or without pop-up questions. Afterwards, students were tested on the entire video content and their test scores compared with an ANCOVA. Interestingly, there was a significant effect of the presence of pop-up questions on these overall test scores after controlling for their exam grade (Fig. 4). Students who watched videos with pop-up questions scored significantly better on the test ( $M_{\text{adj}} = 79\%$ ,  $SE = 1.17$ ) than students who watched the same video without pop-up questions ( $M_{\text{adj}} = 75\%$ ,  $SE = 1.11$ );  $F(1, 218) = 7.68, p = 0.006$  (Online Resource 7). Students with pop-up questions particularly scored more often above 85% when pop-up questions were present (Fig. 4).

### Students' Use of Pop-up Questions

The previous two experiments suggest that pop-up questions do not improve test performance on the specific tested concept, but that merely the presence of pop-up questions affects test performance on the video content as a whole. These results motivated us to perform a set of extra studies and explore possible causes of indirect testing effects. Two semi-structured focus group interviews were performed to investigate how students use pop-up questions. First, students were asked how, where and when they were watching the video. Some students watched the video when commuting in the train or bus but most students watched them at home. Some students explained that they watched the video in one go whereas others said they used their phone or computer at the same time:

**Table 3** Descriptive statistics from an evaluation survey on interactive video for items on video clips and pop-up questions (study 1,  $N = 168$ )

Item	Statement	Likert scale response (%)					N	Mean	S.D
		1 Totally Disagree	2 Disagree	3 Neutral	4 Agree	5 Totally Agree			
3b	The video clips helped me in learning	0 (0%)	0 (0%)	4 (2%)	51 (31%)	108 (66%)	163	4.64	0.53
12a	Answering the pop-up questions helped me in learning	0 (0%)	3 (2%)	11 (7%)	95 (61%)	48 (31%)	157	4.20	0.65
12b	I would like to have fewer pop-up questions within a video clip	38 (24%)	87 (55%)	22 (14%)	9 (6%)	1 (1%)	157	2.03	0.82

The numbers in each category represent the numbers of students answering in that category. The mean (M) and standard deviation (SD) presented are the mean and standard deviation values derived from the Likert scale ranging from 1 (totally disagree) to 5 (totally agree). The sum of the percentages is not equal to 100% due to rounding errors

STUDENT 8: It's not like I am continuously on my phone or something. No. Otherwise I could not follow the story.

STUDENT 9: I sometimes turn that thing (*educational video*) on and then I do something else on my computer in the meantime. (...) I always try to do a bit of multitasking.

Students were then asked to describe their first actions when a question pops up and what they did when their answer to a pop-up question was incorrect. Some students commented that they simply tried the next answer, as they explained:

STUDENT 4: Most often when I receive a question, I just give the answer that I think is right and then I just try the next. It's not like I look back for those things.

A few students specifically clarified that they guessed because they wanted to continue listening to the video lecture:

STUDENT 9: Yes, I do this as quickly as possible because you want to continue the rest of the thing. So you quickly think about it...

Other students explained that they rewind the video when they do not know the answer to a pop-up question, although one

**Table 4** Test scores of students with or without access to corresponding pop-up questions (Study 2,  $N=253$ )

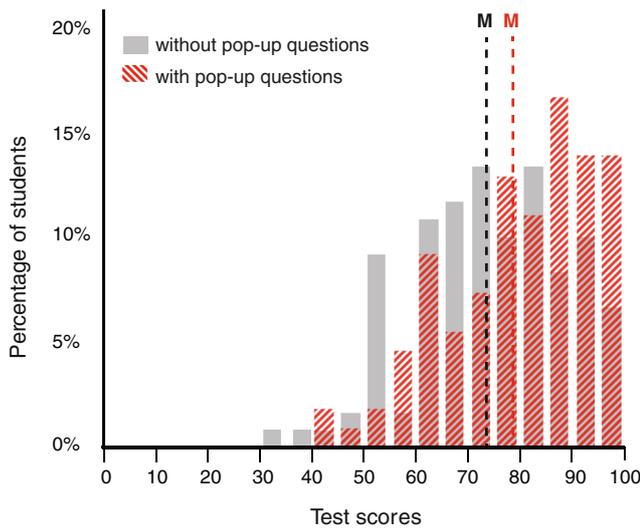
Experimental group	Test Item	Without pop-up question		With pop-up question		<sup>2</sup>	df	p
		N <sup>a</sup>	Correct	N <sup>a</sup>	Correct			
Group A ( $N=151$ )	2	102	44%	97	50%	0.575	1	0.448
	4	85	68%	123	64%	0.359	1	0.549
	6	94	51%	128	54%	0.404	1	0.525
	7	78	72%	103	65%	0.928	1	0.335
	10	81	75%	107	80%	0.694	1	0.405
	12	90	86%	114	86%	0.007	1	0.934
Group B ( $N=108$ )	14	70	81%	89	85%	0.450	1	0.502
	1	148	95%	103	95%	0.038	1	0.846
	3	139	94%	103	98%	2.172	1	0.196 <sup>b</sup>
	5 <sup>c</sup>	103	45%	79	73%	15.098	1	<0.001***
	8	124	90%	89	96%	2.003	1	0.157
	9	121	30%	84	29%	0.033	1	0.855
	11	89	67%	70	66%	0.051	1	0.821
13	109	57%	85	57%	0.003	1	0.954	

The percentages represent the percentage of students with a correct answer to the test questions. <sup>2</sup> and  $p$ -values show the results of a Pearson's chi-square analysis on test scores between the student groups with and without corresponding pop-up questions

<sup>a</sup> The number of participants differs per test item as data were removed for analysis for students not doing the specific test or not watching the corresponding video clip before the test

<sup>b</sup> The  $p$ -value for item 3 was calculated with a Fishers' exact test, because one of the frequencies had an expected count below 5

<sup>c</sup> Test item 5 was nearly identical to the corresponding pop-up question



**Fig. 4** Percentage of students with distinct test scores for groups that watched a video clip on cell signaling with or without pop-up questions (study 3). The possible test scores ranged from 0 to 100. The dashed lines represent the mean adjusted test scores for both the control group ( $N = 83$ ) and the experimental group ( $N = 87$ )

explains that he/she only does this when preparing for final exams and not when preparing for in-class activities:

STUDENT 10: For me it is an indicator of understanding the previous fragment. Usually, when I answer incorrectly, I rewind part of the video.

Similar results were found from the subsequent questionnaire, showing that 47% of all students indicated that they guessed the answer until they found the correct one (Table 5). About 37% of the students indicated rewinding the video first when not knowing the answer to the question. The remaining students claimed to search for the answer on the Internet or in the textbook.

In order to get more insight into the influence of pop-up questions on students’ rewinding behavior, we used learning analytics data. We determined the use of the rewind buttons for both the experimental group with pop-up questions and the control group without pop-up questions. The effect of pop-up questions on rewinding behavior was analyzed from the relative number of rewinds through the course of a video clip (Fig. 5).

These results reveal that students rewind relatively more often within the 30 s after pop-up questions occur ( $M = 0.22$ ,  $SD = 0.23$ ) as compared to the same time points in the control video without questions ( $M = 0.10$ ,  $SD = 0.12$ );  $t(168) = -4.535$ ,  $p < 0.001$ . Similar results were found for a comparable video clip (Online Resource 8).

We also explored the effect of pop-up questions on the general use of both rewind and fast-forward buttons throughout the entire video. Interestingly, students in the experimental group (with pop-up questions) rewound significantly less ( $Mdn = 3$ ) compared to the control group ( $Mdn = 8$ ),  $U = 3946.50$ ,  $p < 0.001$  (Fig. 6a). In addition, students also fast-forwarded significantly less when pop-up questions were present ( $Mdn = 0$ ) as compared to when no pop-up questions appeared ( $Mdn = 5$ ),  $U = 3524.00$ ,  $p < 0.001$  (Fig. 6b). Similar results were found for a comparable video clip (Online Resource 9). No significant difference in the average number of rewinds was found for a control clip similar for both groups. Thus, students rewind and fast-forward less often throughout the video clip as a whole, although they do rewind more often just after pop-up questions appear.

## Discussion

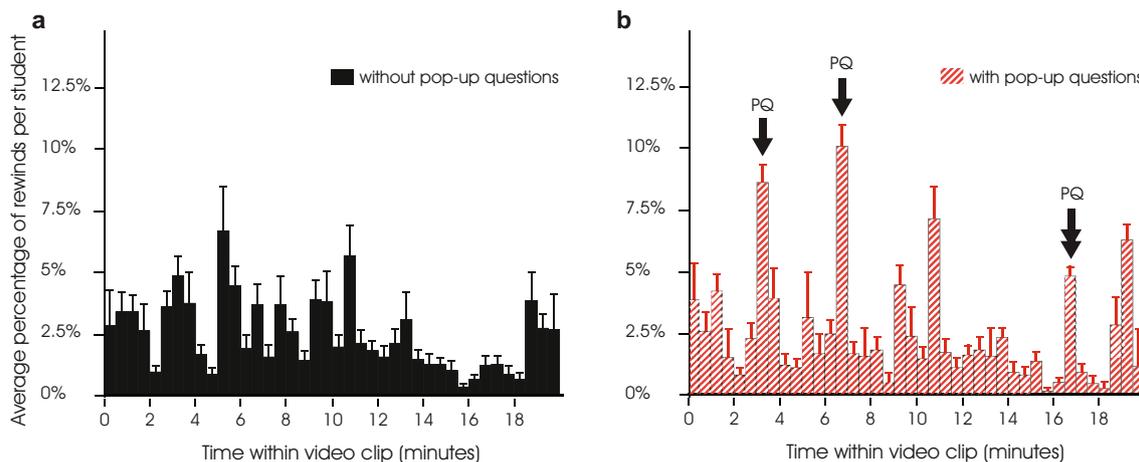
The study demonstrates that pop-up questions within educational videos improve students’ test performance on the overall video content. Accordingly, students agreed that pop-up questions within educational videos helped them to study at home and were positive about including more pop-up questions within the videos. However, pop-up questions on a particular concept within the video did not improve test performance on that specific concept. Thus, our pop-up questions did not result in a direct effect, but rather in an indirect effect on students’ test performance.

It is surprising that we did not find a direct testing effect for pop-up questions, since such an effect has been reported by several previous studies (Butler 2010; Glover 1989; Karpicke and Roediger 2008; Szpunar et al. 2008). One explanation is that studies on the direct testing effect mainly addressed the memorization of vocabulary lists in which students are tested at the level of remembering of

**Table 5** Students’ responses to the statement “What do you do first when you do not know the answer to a pop-up question?” (study 4B,  $N = 118$ )

Item	Statement	Likert scale response (%)			
		Guess answer	Rewind video	Search on Internet	Search in textbook
5	What do you do first when you do not know the answer to a pop-up question?	47%	37%	9%	6%

The sum of the percentages is not equal to 100% due to rounding errors



**Fig. 5** Average percentage of rewinds per student within the time course of a video clip with (a) and without (b) pop-up questions (study 4C). The percentages of rewinds are shown for every 30 s of the video for both the control group ( $N = 117$ ) and the experimental group ( $N = 120$ ). The

arrows illustrate time points of the pop-up questions shown to the experimental group at 3m22s, 6m31s, 16m39s, and 16m57s. Error bars represent standard errors of the means

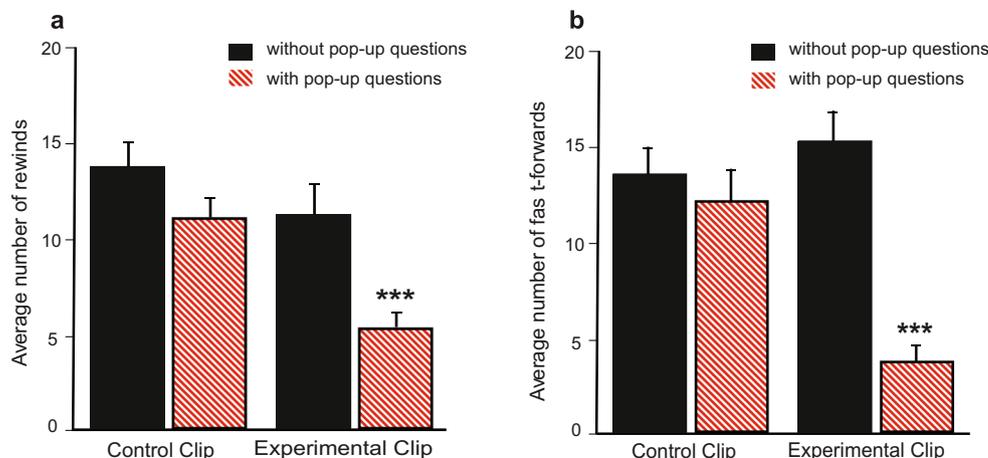
Bloom's taxonomy (Bloom et al. 1956). In the present study, however, students were tested at the level of comprehension. Interestingly, one of the test questions accidentally appeared to be designed for the level of remembering since it was nearly identical to the corresponding pop-up question itself. This pop-up question was also the only question that resulted in a significantly higher score of the corresponding test question. This finding suggests, although speculative, that pop-up questions and answers were remembered but simply not improved students' comprehension of that specific concept.

This study was performed in an authentic setting, meaning that we could only slightly control how students watch the educational videos and how they answer the tests. Therefore, one limitation of this study is that we could not control whether students used any help when performing tests at home. One other limitation of this authentic setting is that data acquisition occurred over multiple years, leading to subtle differences in the course set-up between experiments. Nonetheless,

conclusions were only drawn by comparing results of groups within one cohort. Students were randomly divided over these groups and group results on test performances were corrected for differences in their knowledge.

Future studies are required to determine whether differences in cognitive levels of pop-up questions affect learning differently. It is however unlikely that the use of questions at the level of evaluation will improve the conceptual understanding, since Cummins et al. (2016) reported low study engagement for such pop-up questions. These results were confirmed by the students in our study, who claimed that they would not benefit from more difficult questions, as this would only stimulate them to guess and click through all of the possible answers until correct. Nonetheless, we recommend future studies to investigate different parameters of pop-up questions that might result in direct testing effects such as the level and the frequency of pop-up questions or whether pop-up questions either review or preview the video content.

**Fig. 6** Average number of rewinds (a) and fast-forwards (b) during videos with and without pop-up questions (study 4C). The experimental video clip contained four pop-up questions for the experimental group ( $N = 120$ ) and no pop-up questions for the control group ( $N = 117$ ). The control video clip contained four pop-up questions for both the control group ( $N = 132$ ) and the experimental group ( $N = 127$ ). Error bars represent standard errors of the means



The lack of a direct effect on concept understanding may also be partly explained by the following explorative studies showing that nearly half of the students claimed to guess the answer to a pop-up question and thus did not review the learning content. Some of the students did search for the right answer, either by rewinding or studying other sources, although the effect of this more dedicated approach did not show in the test results. The learning analytics data confirmed that students rewound the video more often just after a pop-up question. A similar effect has been reported before in a study on text instead of videos (Rouet et al. 2001). Rouet et al. provided online texts to students and recorded their scrolling behavior. Interestingly, these students appeared to reread previous information more often when in-text questions were present. The authors of this study propose that text reviewing promotes and guides a deeper level of text comprehension. However, our study did not investigate the effects of the different approaches of students towards pop-up questions. It would be interesting to determine whether students that review the content, do have an increase in conceptual understanding when analyzed separately.

Although we do not report a direct testing effect for pop-up questions, we do show that merely the presence of pop-up questions promoted student performance in the test as a whole. Possible mechanisms for such indirect testing effects have been proposed from previous studies on interpolated tests between video fragments (Lawson et al. 2006; Szpunar et al. 2013; Vural 2013). For example, students have reported to show less mind-wandering when these interpolated tests were present (Szpunar et al. 2013). Although we did not specifically examine mind-wandering, our learning analytics data do reveal that students rewind and fast-forward less over the course of an entire video when pop-up questions are present. We hypothesize that this decrease in zapping back and forth through a video might actually be a result of a higher focus of attention, and this might be particularly true when using relatively long educational videos such as the videos in this study. Other previously reported mechanisms are more note-taking and spending more time on the learning material when video fragments are interpolated with tests (Lawson et al. 2006; Vural 2013). Just the presence of pop-up questions could hence increase students' attention to the video.

In conclusion, our results suggest that teachers can manipulate students' attention and (re-)viewing behavior by inserting pop-up questions within educational videos. Hence, pop-up questions can improve students' learning when watching videos at home. This finding is of particular interest for teachers in a flipped classroom setting who design videos as a preparation for in-class activities.

**Acknowledgments** We gratefully acknowledge the blended learning stimulation program Educate-it of Utrecht University for its support in the application of interactive videos. We thank Rianne Bouwmeester for her comments on the manuscript, Ineke Lam for her contribution to the evaluation survey, and Karin van Look for the focus group interviews.

We thank teachers Dr. Fons Cremers, Dr. Laurens van Meeteren, and Dr. Ron Habets and all students for their participation in this study.

**Author Contributions** All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by M.E. Haagsman, K. Scager, J. Boonstra and M.C. Koster. The first draft of the manuscript was written by Marjolein Haagsman and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Compliance with Ethical Standards

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

**Informed Consent and Statement of Human Rights** All procedures performed in this research, involving human participants, were in accordance with the ethical standards of the institutional and/or national research committee (Review Board of Social Sciences at Utrecht University, IRB approval number FETC180-962). An informed consent was obtained from all individual participants or anonymous data collection was used.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baeppler, 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>.
- Barral, A. M., Ardi-Pastores, V. C., & Simmons, R. E. (2018). Student learning in an accelerated introductory biology course is significantly enhanced by a flipped-learning environment. *CBE—Life Sciences Education*, 17(3), ar38. <https://doi.org/10.1187/cbe.17-07-0129>.
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom : A survey of the research. Resource Document. American Society for Engineering Education. <http://www.asee.org/public/conferences/20/papers/6219/view>
- Bloom, B. S., Englehard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals (handbook I)*. New York: David McKay Company.
- Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE—Life Sciences Education*, 15(4), es6. <https://doi.org/10.1187/CBE.16-03-0125>.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn*, 36(5), 1118–1133. <https://doi.org/10.1037/a0019902>.

- Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2015). *Biology, A global approach (10th inter)*. Harlow: Pearson Education.
- Campbell, N. A., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Reece, J. B. (2017). *Biology, A global approach (11th inter)*. Harlow: Pearson Education.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/0096-3445.135.4.553>.
- Cummins, S., Beresford, A. R., & Rice, A. (2016). Investigating engagement with in-video quiz questions in a programming course. *IEEE Trans Learn Technol*, 9(1), 57–66.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *J Educ Psychol*, 81(3), 392–399. <https://doi.org/10.1037/0022-0663.81.3.392>.
- Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE Life Sciences Education*, 14(4), ar36. <https://doi.org/10.1187/cbe.15-02-0040>.
- HSP [Online application software]. (2020). Retrieved from <https://h5p.org>
- Hapyak Interactive Video [Online application software]. (2020). Retrieved from <https://corp.hapyak.com>
- Herreid, C. F., & Schiller, N. A. (2013). Case studies and the flipped classroom. *J Coll Sci Teach*, 42(5), 62–66.
- Hihaho [Online application software]. (2020). Retrieved from <https://hihaho.com>
- Ibrahim, M. (2011). Implications of designing instructional video using cognitive theory of multimedia learning. *Critical Questions in Education*, 3(2), 83–104.
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *J Exp Psychol Learn Mem Cogn*, 36(6), 1441–1451. <https://doi.org/10.1037/a0020636>.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(February), 772–775.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>.
- Kumar, D. D. (2010). Approaches to interactive video anchors in problem-based science learning. *J Sci Educ Technol*, 19(1), 13–19. <https://doi.org/10.1007/s10956-009-9154-6>.
- Lavigne, E., & Risko, E. F. (2018). Optimizing the use of interpolated tests: The influence of interpolated test lag. *Scholarsh Teach Learn Psychol*, 4(4), 211–221. <https://doi.org/10.1037/stl0000118>.
- Lawson, T. J., Bodle, J. H., Houlette, M. A., & Haubner, R. R. (2006). Guiding questions enhance student learning from educational videos. *Teach Psychol*, 33(1), 31–33. [https://doi.org/10.1207/s15328023top3301\\_7](https://doi.org/10.1207/s15328023top3301_7).
- Lax, N., Morris, J., & Kolber, B. J. (2016). A partial flip classroom exercise in a large introductory general biology course increases performance at multiple levels. *J Biol Educ*, 51(4), 412–426. <https://doi.org/10.1080/00219266.2016.1257503>.
- Lo, C. K., & Hew, K. F. (2017). A critical review of flipped classroom challenges in K-12 education: Possible solutions and recommendations for future research. *Res Pract Technol Enhanc Learn*, 12(1), 4. <https://doi.org/10.1186/s41039-016-0044-2>.
- Mayer, R. E. (2002). Multimedia learning. *Psychol Learn Motiv*, 41, 85–139.
- Panopto™ [Online application software]. (2020). Retrieved from <https://www.panopto.com>
- Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Front Psychol*, 5. <https://doi.org/10.3389/fpsyg.2014.00286>.
- Playposit [Online application software]. (2020). Retrieved from <https://go.playposit.com>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn Sci*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory; basic research and implications for educational practice. *Perspect Psychol Sci*, 1(3), 181–210.
- Rouet, J., Vidal-Abarca, E., & Erbou, A. B. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Process*, 31(2), 163–186. <https://doi.org/10.1207/S15326950DP3102>.
- ScalableLearning [Online application software]. (2020). Retrieved from <https://www.scalable-learning.com/#home>
- 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*. <https://doi.org/10.1037/a0013082>.
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proc Natl Acad Sci*, 110(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>.
- 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>.
- Vural, O. F. (2013). The impact of a question-embedded video-based learning tool on e-learning. *Educational Sciences: Theory Practice*, 13(2), 1315–1323.
- Wieling, M. B., & Hofman, W. H. A. (2010). The impact of online video lecture recordings and automated feedback on student performance. *Computers and Education*, 54(4), 992–998. <https://doi.org/10.1016/j.compedu.2009.10.002>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.