Harnessing active engagement in educational videos: Enhanced visuals and embedded questions

Greg Kestin [1] and Kelly Miller [2]

[1] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[2] *Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 25 March 2021; revised 19 November 2021; accepted 5 April 2022; published 21 June 2022)

The prevalence of online instruction highlights the importance of videos in education. Pedagogies that include elements that actively engage students are accepted as an improvement over more passive modes of instruction. How can we transfer the advantages of active engagement to instruction via video? Previous research on instructional videos has shown that there are a number of principles, the adherence to which benefit student learning by maximizing productive cognitive processing. To understand the impact of combining such principles we designed and produced four different versions of the same physics demonstration video, varying levels of "visual enhancement" designed around these principles and the amount of active engagement across the different versions. Using pre-post video testing, we compared how much viewers learned across the four different versions. We found that actively engaging students by embedding questions throughout the video increases student learning. We also found that physics videos are most effective when they include enhanced visuals *and* embedded questions. Notably, it is the combination that matters most; the learning effect from embedding questions is increased when the video also includes enhanced visuals. This study represents an important step towards understanding how instructors can design and refine their videos to maximize student learning.

DOI: 10.1103/PhysRevPhysEducRes.18.010148

I. INTRODUCTION

Over the past decade there has been increasing interest in teaching and learning online and a trend towards instruction through videos [1]. While instruction through video does not inherently provide pedagogical advantages over live instruction [2], videos can be more effective when particular features are included in their design.

Here we explore instructional videos in the context of physics demonstration videos. When teaching a physical science course, in person or online, it is essential that instructors provide students with opportunities to directly observe the physical phenomena that illustrate the concepts they are trying to teach. Students' observations of a physical demonstration can highlight their misconceptions and increase their interest [3]. When teaching in person, instructors mostly achieve this by relying on live, in-person lecture demonstrations. Online courses instead must rely on videos of demonstrations.

In previous work we have shown that physics demonstration videos can lead to improved learning over nearly identical live demonstrations [4]. These videos can be categorized as *instructional* explanation videos, which are short (approximately 3–4 min) videos with the goal of explaining a specific concept [5]. We found that students who viewed these demonstration videos learned 25%–30% ($p < 0.01$) more than those who viewed the live, in-person demonstration. Other previous work shows that correctly observing the outcome of a demonstration is one of the most important predictors for students successfully learning physics from a live lecture demonstration [6]. One of the reasons why students learn more from demonstration videos (compared to live demos) could be that the outcome of the demo is easier to observe in a video [4,6], due to additional capabilities such as freezing the frame and highlighting a relevant region of the video [7]. Here we present a first step in exploring why students appear to learn more from video demonstrations than live demonstrations and in understanding specifically which features, when included in a video, are most important for helping students learn.

The first step in creating a video is choosing the content and constructing a "narrative"—this essentially boils down to writing a script (i.e., narration).[1] This narration is then paired with appropriate visuals, and then finally, for

[1] This claim draws on nearly a decade of experience of one of the authors (G. K.) writing and producing videos and documentaries (e.g., via NOVA | PBS).

educational videos, interactive elements, such as embedded questions, may be included. While this order is not universal nor the steps neatly delineated, the three components at play are narrative, visuals, and active engagement elements. In the present study, all videos have the same narration, while the visuals and active engagement elements will vary in their level of adherence to research-based practices.

The literature points to a group of principles that, when followed within video design, improve viewers' cognitive processing. Cognitive processing is defined as mental functions involved in the acquisition, storage, interpretation, manipulation, transformation, and use of knowledge [8]. The cognitive theory of multimedia learning points to three design goals for videos to optimize productive cognitive processing of information: reducing extraneous processing,[2] managing essential processing,[3] and fostering generative processing[4] [9–11]. These goals can be achieved by adhering to various principles, including signaling, temporal contiguity, and modality matching [10].

Matching modality is the process of using both audio and visual channels to convey new information, ensuring that the particular type of information fits with the most appropriate channel. Narrating a process while, at the same time, showing an animation of the process uses both channels to explain the process and this gives the learner complementary streams of information. Temporal contiguity is the idea that students learn better when related words and pictures happen simultaneously [5]. Signaling, or highlighting the most relevant parts of an explanation, is a feature shown to promote effective instruction, both within the cognitive theory of multimedia learning, as well as within more recent frameworks for the design of instructional videos [12,13]. A proven way to highlight the relevance of information is embedding visuals (graphical overlays and animations) that signal the essential ideas of an explanation [13]. Signaling has been shown to be important for motivating students to engage in processing the information on a deeper level [14]. Embedding questions in instructional explanation videos can be considered another example of signaling to the viewer the relevancy of the most important information.

While active learning in the college classroom [15–21] as well as active engagement within a multimedia lesson have been shown to improve students' learning [7], we aim to explore how a simple and common type of active engagement, namely, embedded questions, impacts students' learning when included in an instructional video. There is existing research on how the inclusion of questions, independent of variation in visual presentation, improves learning gains [5,7,22].

While we focus on specific ways to increase active engagement, educational psychology literature points to increased cognitive activation as being a core feature of quality instruction generally. Improved learning through inclusion of elements that foster active processing is not unique to video, for example, prompts have been shown to increase learning outcomes in written instructional explanations [23]. Also, while we engage students with discrete questions, microlevel interaction, such as pausing and replaying parts of a video may also increase learning [24]. The capability for microlevel activities such as pausing and replaying does not vary between our conditions and therefore is not a consideration in this study.

We investigate the effect of videos which *simultaneously* actively engage students (by embedded questions) and include enhanced visuals. We refer to videos which employ features such as signaling, temporal contiguity, and modality matching as having "enhanced visuals." We enhance the videos' visuals by carefully coordinating narration, overlaid graphics, text, and on-screen action (see next section for examples). We embed questions throughout the videos, at approximately 30-sec intervals. Using pre-post testing we compare how much participants learn from different versions of the videos, which vary both in terms of visual enhancement and active engagement, to determine the impact of these two elements on learning from videos.

We find that actively engaging students by embedding questions throughout the video improves student learning. We also find that physics videos are most effective when they are designed to both actively engage students *and* have enhanced visuals. Notably, it is the interaction that matters most; the learning effect from embedding questions is enhanced when the video also has enhanced visuals. These results provide useful insights that may help educators design instructional videos to enhance student learning.

II. EXPERIMENTAL METHODS

A. Experimental design

For each of two physics demonstrations, we designed and produced four different versions of the same short (approximately 3–4 min) educational video. The audio track for each video was identical across all four versions. We varied the four versions along two different dimensions: visual enhancement and active engagement. As illustrated in Fig. 1, the videos were designed to explore the effect of these two elements of interest. One of the versions (version 1) was designed both to have enhanced visuals and to actively engage the viewer (through embedded questions). Two of the versions included only one of these elements, embedded questions only (version 2) or

[2]Cognitive processing that does not serve the instructional objective and is caused by poor instructional design.

[3]Cognitive processing required to mentally represent the essential material in the lesson as presented and is caused by the complexity of the to-be-learned material.

[4]Cognitive processing required to make sense of the essential material and depends on the learner's motivation to exert effort during learning.

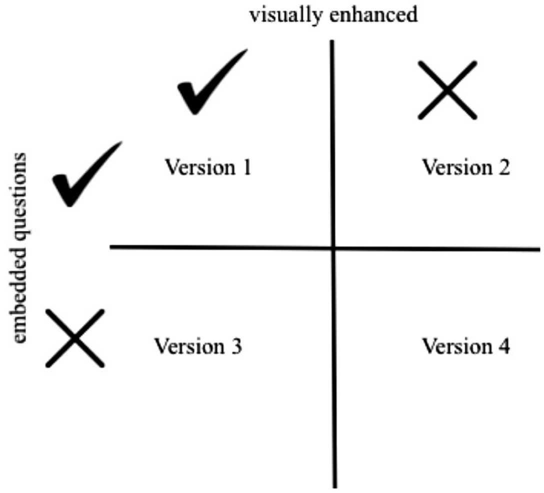

FIG. 1. Summary of four versions of videos used in the study.

enhanced visuals only (version 3). Version 4 included neither enhanced visuals nor embedded questions.

B. Physics demonstration videos

We created the four video versions described above for two different physics demonstrations, "shoot the monkey" and "high road, low road." In the shoot the monkey demonstration a cannon is aimed directly at a (stuffed) monkey suspended a few meters above the ground, at a distance of several meters away from the cannon (see Fig. 2). The cannon is fired and, simultaneously, the monkey is released from rest. As the acceleration due to gravity is the same for both the monkey and the bullet, gravity displaces both objects the same vertical distance in the time that it takes the bullet to travel horizontally to the monkey and the two objects collide (i.e., the bullet hits the monkey). This is used to demonstrate the independence of the horizontal and vertical components of the motion of projectiles.

In all four versions of the shoot the monkey video, the narrator discusses three contrasting cases of the demo: (i) aiming *above* the monkey, (ii) aiming *below* the monkey, and (iii) aiming *directly at* the monkey. In the versions designed to have enhanced visuals (versions 1 and 3), the contrasting case narration is accompanied by animated visuals (i.e., red lines placed at different angles, as seen in Fig. 3) which illustrate the differences between the cases. In the other two versions of the shoot the monkey video (versions 2 and 4), these contrasting cases were described in the narration but were not matched with complementary graphics or animations.

Figure 4 further illustrates the difference between the shoot the monkey demo video versions designed to have enhanced visuals (1 and 3) compared to the versions which were not designed to have them (2 and 4). Figure 4(a)

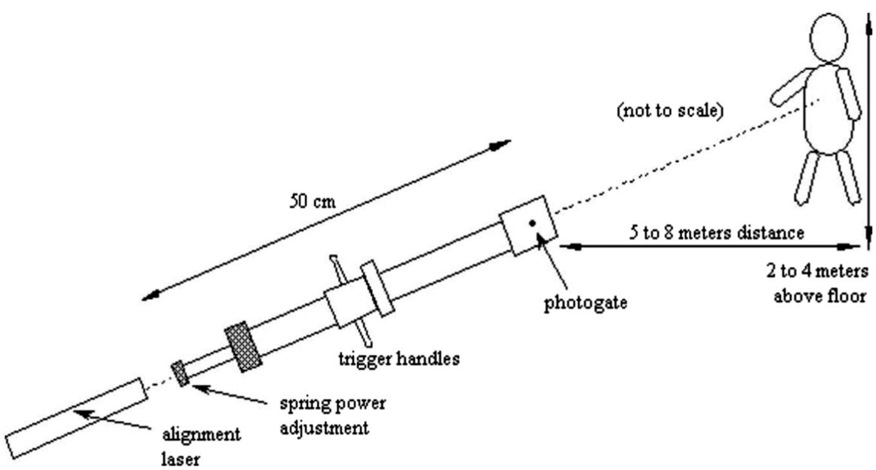

FIG. 2. Shoot the monkey setup schematic.

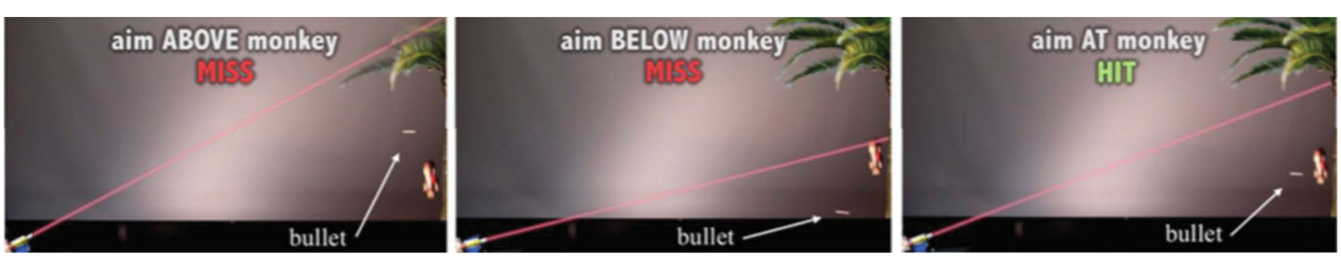

FIG. 3. Screenshots from the shoot the monkey demo video versions which had enhanced visuals (versions 1 and 3).

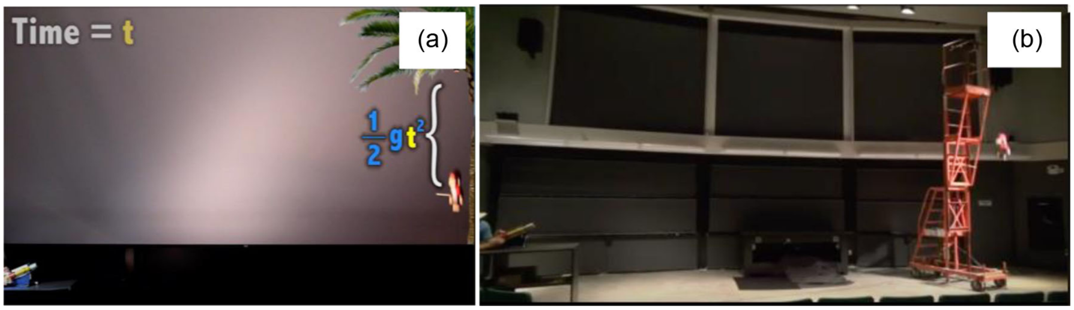

FIG. 4. Screenshot from the shoot the monkey demo video versions which (a) had enhanced visuals (versions 1 and 3) and (b) did not have enhanced visuals (versions 2 and 4).

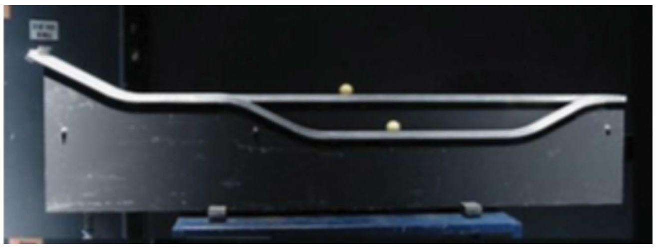

FIG. 5. Images of two tracks in the high road, low road demonstration.

shows a screenshot from video versions 1 and 3 where the video is overlaid with a graphic that signals the important information. In these two versions, a graphic is used to illustrate that since both the bullet and the monkey begin their descent at the same time ($t = 0$) and they are both falling for the same amount of time, they both have the same vertical displacement due to gravity: half the gravitational acceleration (g) times the time elapsed squared. Figure 4(b) shows a screenshot from video versions 2 and 4 which do not include any of these enhanced visuals.

In the high road, low road demonstration, two balls, starting with the same initial horizontal velocity, roll along two different tracks. One ball follows a straight, horizontal track (high road) while the other rolls down a track shaped like a valley with a flat section at the bottom, and then rolls back up to the original height (low road). The horizontal distance traveled by each ball is the same, but the low road track is a longer path. Figure 5 shows a picture of the two tracks.

When the ball on the low road travels into the valley, the x component of its velocity is always larger than that of the ball on the high road and therefore the low road ball "wins the race" despite traveling a longer path. Figure 6 illustrates the difference between the high road, low road demo video versions with enhanced visuals (1 and 3) compared to the versions without enhanced visuals (2 and 4). Figure 6(a) shows a screenshot from video versions 1 and 3 where the video is overlaid with a graphic showing the normal force and weight of the ball; these forces are visually displayed throughout the ball's motion. These enhanced visuals illustrate that the normal force has an x component along one path but not along the other. This helps clarify why the low-road ball accelerates horizontally compared to the high-road ball which moves at a constant speed. These graphics are accompanied by an explanation of the physics which also serves to help manage the viewers' essential processing through modality matching between the auditory and visual channels. The auditory and visual channels describe the same concept adhering to the principle of temporal contiguity. Figure 6(b) shows a screenshot from the corresponding video versions (2 and 4) which do not have enhanced visuals. The physics behind these graphics is explained in the narration (the audio is the same across all four video versions) but, as the visuals are missing, there is no modality matching in these versions of the videos.

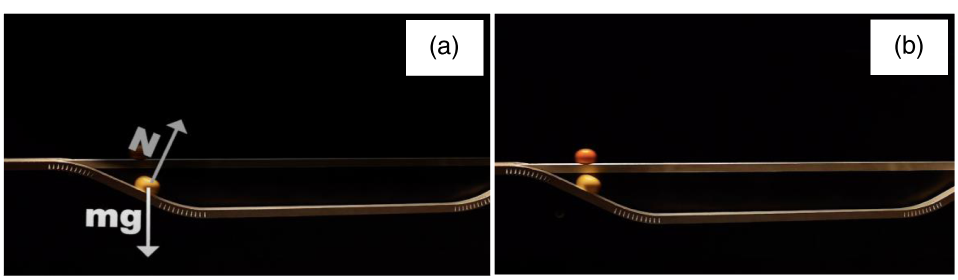

FIG. 6. Screenshot from the high road, low road demo video versions which (a) had enhanced visuals (versions 1 and 3) and (b) did not have enhanced visuals (versions 2 and 4).

C. Embedded questions

For both demonstrations, video versions 1 and 2 included embedded questions. The questions are designed to improve learning in four ways: (i) they actively engage the viewer, (ii) they segment the videos, (iii) they prime the viewer for the material in the subsequent video segments [25], and (iv) they highlight, or signal, the most important information. Each video was broken up into short (approximately 30-sec) segments, and one multiple-choice question was embedded after each segment. Viewers were required to respond to the question before being able to proceed to the next segment of the video. The shoot the monkey demo video included six embedded questions, while the high road, low road demo included four embedded questions.

D. Pretests and post-tests

Each version of each video was embedded in a Google form which included a prevideo test and a postvideo test. Figure 7 summarizes this design.

The prevideo test included four conceptual physics questions which were topically relevant to the content of the video and used as a baseline for conceptual physics understanding. The pretest also included a multiple-choice background question, which asked participants to indicate the highest-level physics course they had completed. The choices were: "I have never taken physics before," "I have completed high school physics," "I have completed college level physics," or "I have completed graduate school level physics." The postvideo test included 4–7 conceptual physics questions which were different but isomorphic to those in the prevideo test. All pre- and post-tests were rescaled to a 5-point scale (so the data from each demo could be pooled). The pre- and post-tests for all four versions of each demo video were identical. The pretests and post-tests for each video are included in Appendices A and B in the Supplemental Material [26].

E. Mechanical turk

We published each of the four versions of the Google form for each demo video on the Amazon Mechanical Turk (MTurk) platform as a Human Intelligence Task (HIT). MTurk is a marketplace for human tasks (HITs) such as completing surveys or generating training data for machine learning. Customers who complete HITs are called "workers" and customers publishing the HITs are called "requesters." Workers are paid for completing HITs (on average $6/hour) and requesters can specify which workers are eligible to complete their HITs. Requesters can restrict workers based on location, level of education, or other demographic criteria. Requesters can also specify that only "masters workers" are eligible to complete their HITs. Masters workers have consistently demonstrated a high degree of success in performing a wide range of HITs across a large number of requesters. Prior research has shown that crowdsourcing tools like MTurk provide subjects (i.e., the workers) which serve as an appropriate substitute for "real" students [27].

FIG. 7. Experimental design.

F. Sample

Participants in this study were MTurk masters workers who completed high school in the United States. Workers were randomly assigned one of the four versions of the video or Google form. We filtered workers' responses on two measures to ensure that we only included, in our final sample, participants who took the survey seriously and answered the questions earnestly. We dropped from the sample any workers who completed the task too quickly (less time than the length of the video plus 2 min). We also included, in the post-test, a "filter" question which anyone reading the questions carefully should have been able to answer correctly. The question asked the participant to imagine there were two cars traveling, one twice as fast as the other. The participant was asked to determine which car would travel further, in the same amount of time. We

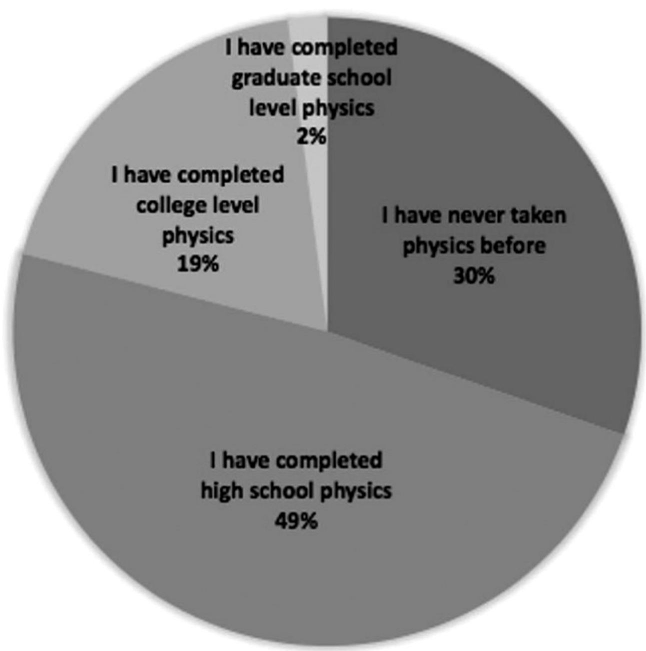

FIG. 8. Level of physics background of Mechanical Turk participants in the study ($N = 300$).

assumed that any participant who failed to answer this question correctly was not reading the questions carefully and we dropped them from the study. We dropped 11 participants who completed the survey too quickly and 15 participants who failed to answer our filter question correctly. Overall, not counting those we dropped, 300 MTurk workers participated in the experiment; between 60 and 80 unique workers completed each of the four versions of the form. Workers were prevented from participating in the study more than once (i.e., each worker completed only one version of the form). The four versions of the survey were posted to Mechanical Turk all at the same time and, from the MTurk worker point of view, they were all the same HIT. There was no systematic assignment of participants to version and therefore, participants were randomly assigned to one of the four experimental conditions. Most of the workers (79%) had either never taken physics before or had completed high school physics. Figure 8 shows the distribution of workers' physics backgrounds ($N = 300$).

III. RESULTS AND DISCUSSION

Figure 9 shows aggregate average postvideo test scores from the four versions of the demo videos. While the averages presented in Fig. 9 represent results pooled from both demos, each was first analyzed separately; we combined the results only after establishing identical trends for each demonstration. As seen in Fig. 9, participants who watched the video version with both enhanced visuals and embedded questions did significantly better on the postvideo test than participants who watched any of the other three versions (p value of 0.001–0.05). There is a statistically significant difference between version 1 and version 2 ($p < 0.05$), between version 1 and version 3 ($p < 0.001$) and between version 1 and version 4 ($p < 0.01$). The differences between the other versions (between 2 and 3; 2 and 4; 3 and 4) were not significant. By comparing postvideo results of participants who watched version 2 to the pooled results of the participants who watched the other three versions, we calculated the effect size (Cohen's d) [28] of the combination of enhanced visuals and embedded questions to be 0.58 (95% confidence interval 0.32–0.85). According to the standard benchmarks suggested by Cohen, a medium effect size is around 0.5 and a large effect size is 0.8 [29]. We also measured the effect sizes across all four groups using post-estimation (eta-squared) of an ANOVA of postvideo test scores. The effect size across all four groups was 0.07 ($df = 3$, 95% CI = 0.02–0.13). To ensure that the groups of participants across all four versions of the videos were comparable, we did a one-way analysis of variance of prevideo test scores across the four groups of MTurkers and determined that there was no statistically significant difference between any of the four groups ($F = 0.61$, $p = 0.61$). We verified that there was no significant difference in prevideo test scores across the four groups with a Bonferroni test. We also analyzed participants' self-reported "highest level of physics course completed" (the data depicted in Fig. 8) by conducting a one-way analysis of variance. The results of this ANOVA (and a subsequent Bonferroni test) determined that there was no statistically significant difference in the distribution of the "highest level of physics course completed" between the four groups ($F = 1.17$, $p = 0.32$).

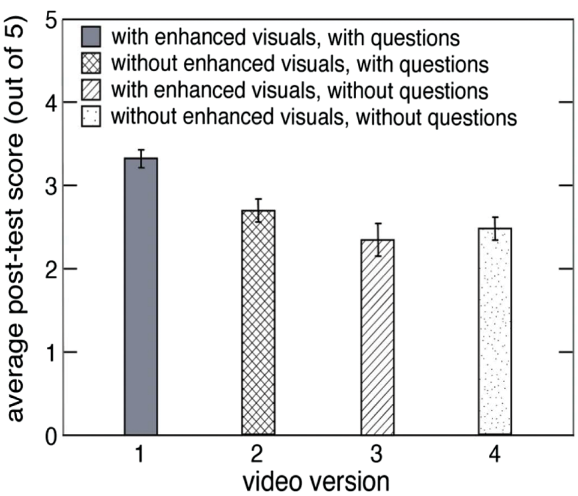

FIG. 9. Average postvideo test score for participants who watched each of the four versions of the videos: version 1: with enhanced visuals and embedded questions; version 2: without enhanced visuals, with embedded questions; version 3: with enhanced visuals, without embedded questions; version 4: without enhanced visuals, without embedded questions. Error bars represent the standard error of the mean.

Figure 9 shows that videos are most effective when they are designed *both* to actively engage students (through embedded questions) and have enhanced visuals. Notably, it is the interaction that matters most; videos that either solely promote active engagement or solely include enhanced visuals are not as effective as videos that do both. Table I shows the standardized coefficients for two linear regression models predicting the postvideo test scores with the two variables of interest (visual enhancement and embedded questions). Both models control for participants' levels of physics background and their prevideo test scores. The level of physics background is controlled for by including participants' self-reported highest level of physics course completed as a categorical variable in each model. There is only a weak correlation between the highest level of physics course a participant completed and their prevideo test scores (0.11, $p = 0.04$) and we included both as control variables, as doing so improved the predictive power of both models. Model 1 shows the main effects of embedded questions and enhanced visuals on postvideo test scores. Embedding questions in the videos is predictive of higher postvideo test performance ($p < 0.01$) but the effect of enhanced visuals (on their own) is not predictive of participants'

TABLE I. Standardized coefficients for linear regression models predicting students' postvideo performance predicted by the main effects of enhanced visuals and embedded questions (model 1) and the interaction between enhanced visuals and embedded questions (model 2) (both models control for performance on prevideo questions and incoming physics background).

	Model 1	Model 2
	Standardized coefficients	
Constant	−0.35	−0.24
Performance on prevideo questions	0.17^{**}	0.16^{**}
Level of physics background	0.08	0.08
Enhanced visuals	0.19	−0.05
Embedded questions	0.36^{**}	0.14
Interaction between enhanced visuals and embedded questions		0.47^{*}
R^2	0.06	0.07
RMSE	0.97	0.96

$^{***}p < 0.001$
$^{**}p < 0.01$
$^{*}p < 0.0$

postvideo test scores. This finding is in keeping with the literature—students learn more from videos when they are required to answer relevant questions [5]. The interpretation of this increased learning from embedded questions has a subtlety, which is illustrated by model 2. Model 2 shows that the interaction between the two variables of interest is significant ($p < 0.05$). This interaction is illustrated in Fig. 10. The difference in postvideo test performance between videos with embedded questions, compared to videos without embedded questions, is increased significantly when embedded questions are coupled with enhanced visuals.

It appears that to maximize learning from active engagement in videos, interactive questions and enhanced visuals need to go hand in hand. In versions 1, we designed each video segment to proceed in two steps: (i) students are presented with a question that actively engages them and primes them for upcoming information presented in the video, (ii) information relevant to the question is presented with enhanced visuals. We should therefore expect that the majority of learning in each video segment occurs not during the period in which the participant is answering the embedded questions, but rather when watching the following segment. In order to test this, we looked at the relationship between performance on the embedded questions and postvideo test performance.

Table II looks specifically at the videos with questions embedded (versions 1 and 2) and shows a linear regression model that predicts students' post-test performance based on how they answered the questions embedded throughout the video (controlling for inclusion of enhanced visuals and performance on prevideo questions). The coefficient for performance on embedded questions is not statistically significant, indicating that there is no relationship between answering the embedded questions correctly and performance on the postvideo test. This is consistent with the two-step design of embedded questions described above. That is, embedded questions first serve to prime the viewer for observation of key points in the video that are then highlighted, or signaled, through enhanced visuals. If the ordering of the elements were reversed (video segment signals important information and *then* a relevant question is embedded *after* the video segment, as opposed to *before* the segment) we might expect a testing effect [30]—that is, performance on the embedded questions in the video would be correlated with postvideo test performance. The fact that we see no such correlation suggests that the embedded questions are in fact priming viewers for the video segment which follows each question, and there is no testing effect.

From the literature we would expect that both enhanced visuals (which help foster productive cognitive processing) [10] and active engagement [16–21] in videos would increase viewers' learning. Given that the sequence in which questions and related information relies on priming,

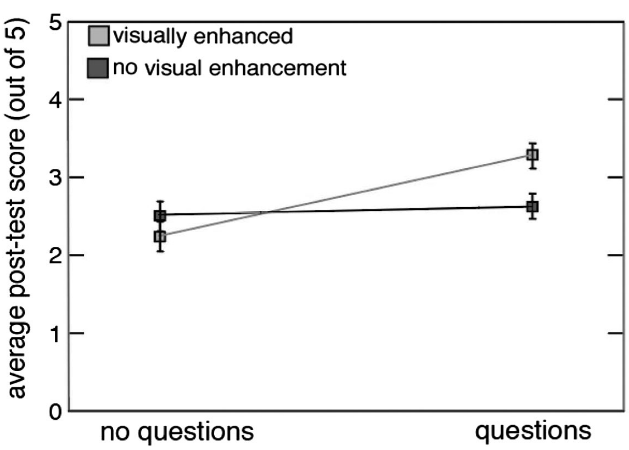

FIG. 10. Model 2 interaction between enhanced visuals and embedded questions in predicting postvideo test scores.

TABLE II. Standardized coefficients for linear regression models predicting students' postvideo performance based on their performance on the questions embedded throughout the video (controlling for performance on prevideo questions and inclusion of enhanced visuals).

	Standardized coefficients
Constant	2.52
Performance on prevideo questions	0.06
Performance on questions embedded in videos	0.06
Enhanced visuals	0.56^{**}
R^2	0.07
RMSE	1.16

$^{***}p < 0.001$
$^{**}p < 0.01$
$^{*}p < 0.05$

we would also expect no correlation between performance on embedded questions and post-test questions. Both of these expectations are realized by the data. The model presented in Table II also confirms that having enhanced visuals is a significant predictor for postvideo test performance, only assuming the videos already have embedded questions.

Finally, while prior knowledge (pretest score) is predictive of absolute post-test score, we find no relationship between prior knowledge and learning *gains*. That is, participants with different levels of prior knowledge appear to benefit equally from videos which have enhanced visuals and include embedded questions. Previous research has shown that students who have a better understanding of the underlying concepts are more likely to observe and remember an in-person demonstration correctly [5]. We suggest that by presenting demonstrations through videos that have enhanced visuals, one fosters correct observation, therefore allowing those coming to the video with a weaker understanding a unique opportunity to learn.

We acknowledge that this study has some limitations. We have performed an experiment with two specific demonstration videos and, while we have found the same trends in our results for both videos, we cannot assume that these results are generalizable to all videos in all subject domains. Also, further work should explore the role of segmenting (dividing the video into segments), which based on the literature, improves learning in some cases but not others [31]. This study also has limitations associated with the sample population. We have performed this experiment with Mechanical Turk workers, not students and not in the context of an academic class. While there is research that shows MTurk workers are an appropriate substitute for students, participating in a platform like the Mechanical Turk might introduce unmeasured, implicit biases in the sample such as a positive selection of participants who are "tech savvy." Further research needs to be conducted to show that these results are generalizable to students in a classroom setting.

While the present investigation focuses on *learning gains* from physics demonstration videos, not all demonstrations are presented with the aim of improving conceptual understanding. From informal surveys of several physics instructors and demonstration technicians, we have found that other aims include: presenting an everyday application of a concept, increasing motivation, and increasing excitement. It is no surprise that demonstrations are often the most memorable part of a physics course. Future work should explore a wider array of the benefits of demonstrations.

IV. CONCLUSION

We have shown that embedding questions in physics demonstration videos improves learning. We have also shown learning gains from videos that include embedded questions and enhanced visuals are higher than those from videos with one or neither of these elements. With the recent increase in online education, this study represents an important step towards understanding how instructors can design demonstration videos to maximize student learning. We recommend that, when creating instructional videos, educators, at the very least, embed questions to prime viewers. Ideally, instructors should also incorporate enhanced visuals, as it appears to be the combination which maximizes learning from videos.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge contributions from and valuable discussions with Louis Deslauriers, Logan McCarty, Kristina Callahan, Eric Mazur, Melissa Franklin, Anna Klales, Wolfgang Rueckner, Allen Crockett, Daniel Davis, and Daniel Rosenberg. This work was supported by Harvard University through the Division of Science in the Faculty of Arts and Sciences and through the Harvard John A. Paulson School of Engineering and Applied Sciences.

[1] E. Cruse, Using educational video in the classroom: Theory, research and practice, Library Video Company **12**, 56 (2006), https://www.safarimontage.com/pdfs/training/usingeducationalvideointheclassroom.pdf.

[2] R. M. Bernard, P. C. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, P. A. Wallet, M. Fiset, and B. Huang, How does distance education compare with classroom instruction? A meta-analysis of the empirical literature, Rev. Educ. Res. **74**, 379 (2004).

[3] J. B. Johnston, The lecture demonstration: A developing crisis, Phys. Teach. **19**, 393 (1981).

[4] G. Kestin, K. Miller, L. S. McCarty, K. Callaghan, and L. Deslauriers, Comparing the effectiveness of online versus live lecture demonstrations, Phys. Rev. Phys. Educ. Res. **16**, 013101 (2020).

[5] C. Kulgemeyer, A framework of effective science explanation videos informed by criteria for instructional explanations, Res. Sci. Educ. **50**, 2441 (2020).

[6] K. Miller, N. Lasry, K. Chu, and E. Mazur, Role of physics lecture demonstrations in conceptual learning, Phys. Rev. ST Phys. Educ. Res. **9**, 020113 (2013).

[7] C. J. Brame, Effective educational videos: Principles and guidelines for maximizing student learning from video content, CBE Life Sci Educ **15**, 1 (2016).

[8] https://dictionary.apa.org/cognitive-process.

[9] R. E. Mayer, Multimedia learning and games, *Computer Games and Instruction* (Information Age Publishing, Charlotte, NC, 2011), pp. 281–305.

[10] R. E. Mayer, Evidence-based principles for how to design effective instructional videos, J. Appl. Res. Mem. Cogn. **10**, 229 (2021).

[11] R. Mayer, *The Cambridge Handbook of Multimedia Learning*, 2nd ed. (Cambridge University Press, New York, 2014).

[12] R. E. Mayer, G. T. Dow, and S. Mayer, Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds?, J. Educ. Psychol. **95**, 806 (2003).

[13] J. Roelle, K. Berthold, and A. Renkl, Two instructional aids to optimise processing and learning from instructional explanations, Instr. Sci. **42**, 207 (2014).

[14] S. R. Acuña, H. G. Rodicio, and E. Sánchez. Fostering active processing of instructional explanations of learners with high and low prior knowledge, Eur. J. Psychol. Educ. **26**, 435 (2011).

[15] R. R. Hake, Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[16] C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[17] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, Science **332**, 862 (2011).

[18] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).

[19] J. M. Fraser, A. L. Timan, K. Miller, J. E. Dowd, L. Tucker, and E. Mazur, Teaching and physics education research: Bridging the gap, Rep. Prog. Phys. **77**, 032401 (2014).

[20] L. Deslauriers and C. Wieman, Learning and retention of quantum concepts with different teaching methods, Phys. Rev. ST Phys. Educ. Res. **7**, 010101 (2011).

[21] L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom, Proc. Natl. Acad. Sci. U.S.A. **116**, 19251 (2019).

[22] M. Keithly *et al.*, Blending it up: Active learning in a STEM classroom through the use of on-line materials, Abstracts of Papers Am. Chem. Soc. **250**, 1155 (2015).

[23] K. Berthold and A. Renkl, How to foster active processing of explanations in instructional communication, Educ. Psychol. Rev. **22**, 25 (2010).

[24] M. Merkt, S. Weigand, A. Heier, and S. Schwan, Learning with videos vs. learning with print: The role of interactive features, Learn. Instr. **21**, 687 (2011), https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.3414&rep=rep1&type=pdf.

[25] E. Tulving and D. L. Schacter, Priming and human memory systems, Science **247**, 301 (1990).

[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.18.010148 for Appendix A and Appendix B, which contain the pretest and post-test for the shoot the monkey and high road low road demonstrations, respectively.

[27] P. M. Sadler *et al.*, Identifying promising items: The use of crowdsourcing in the development of assessment instruments, Educ. Assess. **21**, 1996 (2016).

[28] J. Cohen, Some statistical issues in psychological research, in *Handbook of Clinical Psychology*, edited by B. B. Wolman (McGraw-Hill, New York, NY, 1965), pp. 95–121.

[29] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge Academic, New York, NY, 1988).

[30] C. L. Bae, D. J. Therriault, and J. L. Redifer, Investigating the testing effect: Retrieval as a characteristic of effective study strategies, Learn. Instr. **60**, 206 (2019).

[31] P. Thompson *et al.*, The effect of designing and segmenting instructional video, J. Inf. Technol. Educ.: Res. **20**, 173 (2021), https://jite.org/documents/Vol20/JITE-Rv20p173-200Thompson6921.pdf.