


RESEARCH

Open Access



Learning from watching dialog and monolog videos in online STEM courses

Yingxiao Qian^{1*} , Yi-Chun Hong² and Michelene Chi²

Abstract

Background Instructional videos have been increasingly critical to afford effective learning experiences in online courses, but most videos follow a lecturing monolog delivery format. This format tends to lead students to observe the videos passively. Previous laboratory studies have indicated that observing dialog videos of an instructor tutoring a tutee enhances student learning more than watching monolog videos of the same instructor lecturing. This paper describes two empirical studies that replicated the laboratory findings in large-scale college-level online STEM courses.

Results The results show that observing dialog videos of tutoring led to superior learning outcomes compared to monolog videos of lecturing, regardless of whether students observe the videos individually or collaboratively, as long as they engaged generatively with the materials. This finding was confirmed across two tested STEM domains.

Conclusions The findings of these classroom studies suggest that observing dialog tutoring videos is a novel and robust online instructional format that can be generalized across STEM domains. The benefit of overhearing tutorial dialogs requires students to engage in *Constructive* and *Interactive* behaviors, as defined by the ICAP framework, rather than exposing students to long didactic lectures.

Keywords Instructional video, ICAP framework, Tutoring dialog, STEM education, Online

Introduction

Online courses have become a cost-effective, flexible alternative delivery format for college-level courses, especially during the outbreak of the COVID-19 pandemic. The National Center for Education Statistics (2022) reported that, in the post-pandemic era, there was a continuous increase in college students' online course enrollment worldwide. In the United States, over 9.4 million college students, around 61% of undergraduate student populations, had taken an online course by fall 2021, 28% of whom took courses solely online. Therefore, allowing

college students to maintain an effective learning experience in online courses has become critical.

For a majority of online courses, content delivery mainly relies on videos that students can view, pause, and replay to learn and reflect upon (Auerbach & Andrews, 2018; Barlow & Brown, 2020; Weinberg et al., 2022). Despite flexible options for students to navigate the content, videos are predominantly delivered through a monolog format in online courses. Monolog videos refer to a lecture format videos in which instructors teach-by-telling (Hew & Cheung, 2014; Roscoe & Chi, 2007). This didactic format of videos can provide students with course content knowledge, but its one-way information delivery actually leads to minimal student learning, as predicted by the ICAP framework (Chi, 2009; Chi & Wylie, 2014).

The ICAP theory explicitly operationalizes and differentiates students' engagement activities into a framework of four modes with operational definitions—*Passive*,

*Correspondence:

Yingxiao Qian
yingxiao@mailbox.sc.edu

¹ Department of Leadership, Learning Design, and Inquiry, College of Education, University of South Carolina, Columbia, SC 29208, USA

² Mary Lou Fulton Teachers College, Arizona State University, Tempe, AZ 85281, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Active, *Constructive*, and *Interactive*—based on the overt behavioral activities that students can undertake with the learning materials and the content of the outputs they produce. As defined in Chi and Wylie (2014), the ICAP framework considers *Passive* as the lowest mode of engagement in which students do *paying attention* to the instructional materials, but they do not undertake any other physical interactions with the instructional materials. That is, students produce no outputs in the *Passive* mode. In contrast, the *Active* mode is when students undertake activities that merely *manipulate* the information in the instructional materials without adding new information, such as underlining text sentences. The *Constructive* mode describes students *generating* new knowledge by inferring beyond what was presented in the instructional materials through activities such as drawing diagrams, explaining, asking questions and deriving solutions. The last *Interactive* mode denotes when two or more peers are *co-generating* knowledge collaboratively while dialoguing, such as asking and answering each other's questions or elaborating upon each other's comments. The ICAP framework hypothesizes that the collaborative (*Interactive*) mode has the potential to achieve the highest level of learning (if collaboration is carried out co-generatively), followed next by the generative (*Constructive*) mode, then the manipulative (*Active*) mode, and finally the attentive (*Passive*) mode. The predicted decreasing levels of learning (I > C > A > P) are supported by hundreds of published studies in the literature (Chi, 2009; Chi & Wylie, 2014; Chi et al., 2008). Thus, to achieve a higher level of learning in online courses, it is desirable to elicit more *Constructive*, or *Interactive* student engagement, but how?

In a one-to-one context (either online or in-person context), one-to-one tutoring has been viewed as one of the most effective formats of instruction (Merrill et al., 1995; VanLehn, 2011). However, the high cost of affording one-to-one tutoring sessions makes it unlikely for deployment in large-scale STEM courses. Chi et al. (2008) designed a new paradigm and showed that students who collaboratively observed tutorial dialog videos between a tutor (analogous to an instructor) and a tutee (analogous to a student) can learn just as well as the tutee in the video who was being tutored one-to-one. Moreover, observing tutorial dialogs collaboratively was more effective for learning than a number of control conditions, such as students observing tutorial dialog videos individually (Chi et al., 2008; Muldner et al., 2011; Muller et al., 2008), and beneficial for STEM content domains, such as physics (Chi et al., 2008), mathematics (Lobato et al., 2023), and biology (Muldner et al., 2014). Thus, adopting tutorial dialog videos as a source of instructional videos seems to be a promising new paradigm for online courses.

However, it is important to test those laboratory findings in an “authentic” online learning environments rather than “in-vivo” classroom studies to see if they remain valid. By “authentic” we mean that the instructional materials (including the pretests and post-tests) are created and delivered by the actual instructors, as opposed to “in-vivo” studies, in which the materials and implementation are carefully crafted by the experimenters. Authentic learning environments typically also have many factors that vary from laboratory studies. For example, students in the Chi et al. (2008) laboratory studies watched videos in an in-person setting where their learning processes were observed and videotaped. In addition, the videos were supplemental learning materials (e.g., as an intervention lasting for 7–13 min focused on one topic), whereas in online authentic courses, they could serve as the primary delivery format of the entire course content. Thus, due to the various differences between a laboratory setting and an authentic classroom, it is often the case that robust laboratory findings only replicates in in-vivo studies but do not replicate in authentic classroom or online contexts. For example, although Booth et al. (2013) replicated the advantage of students studying incorrect problem examples rather than correct examples in an in-vivo context of implementation, this finding was not replicated in a more authentic classroom context (Barbieri & Booth, 2016).

Therefore, the first study described below examined whether learning from observing tutorial dialog videos was better than learning from observing lecturing monolog videos, collaboratively and individually, in large-scale online biology courses. Moreover, because implementing collaborative watching of videos in large online classes is often cumbersome, it is important to know whether watching dialog videos individually is more beneficial than watching monolog videos for college students across various STEM domains. Thus, the second study described below examined whether it is equally beneficial for students to watch dialog than monolog videos individually in an authentic online mathematics course.

In short, the two research questions (RQ) corresponding to the two studies were:

RQ1 for Study 1: Can college students learn better from observing dialog videos than observing monolog videos when they watch individually and collaboratively in online courses?

RQ2 for Study 2: Does the learning benefit of observing dialog videos over monolog videos replicate for students when watching individually in a different STEM domain?

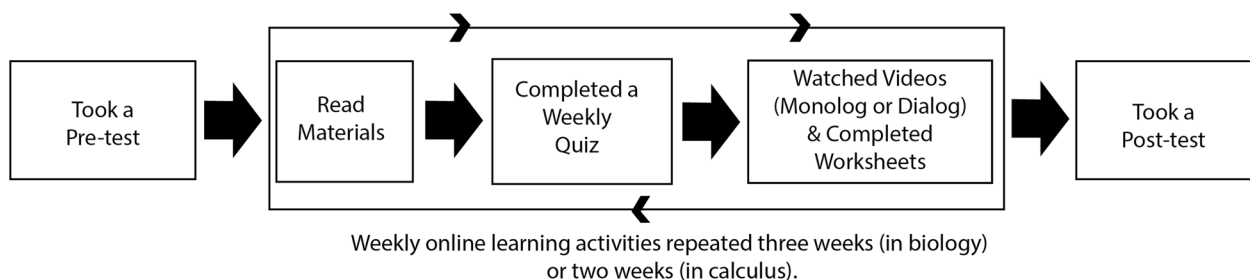


Fig. 1 Sequence of steps all students take

Overview of the two studies

In this section, we focus on the common aspects of both Study 1 and Study 2, described below.

Design of the studies and participants

Both Study 1 and Study 2 used a between-subjects pre-test–post-test design (Gall et al., 2007) by asking college students enrolled in large online STEM courses to observe two different types of videos (e.g., monolog and dialog) either individually or collaboratively. Specifically, in Study 1, we compared whether college students enrolled in a biology class learned better by watching dialog videos than observing monolog videos, either individually or collaboratively, in which different students participated in the four conditions, but watched the same videos, all taught by the same instructor. In Study 2, we compared students learning from observing dialog and monolog videos individually only, in the context of an online mathematics course.

The students were all recruited from online STEM courses at a public university in the southwestern United States. A stratified random sampling method (Gall et al., 2007) was applied to assign students to the conditions, in order to mitigate any pre-existing differences in prior knowledge between the observing conditions (Makwana et al., 2023; Muldner et al., 2014). Specifically, strata were determined based on participants' pretest scores, and then for each stratum, students were randomly assigned to each observing condition.

Procedure

Each study was implemented with six lessons, with each lesson covering one topic taking an entire lesson period. Each week has two (biology) or three (mathematics) lessons, following the same procedure. Pretests were administered at the beginning of the study to assign students to each condition. At the end of each study, all students were asked to complete the post-tests as a course exam. Students in each group accessed assigned materials in a learning management system, Canvas.

For each week, students from each condition were required to complete four activities in a fixed sequence (see Fig. 1 for details). First, the students were asked to read the assigned materials and take a weekly quiz to assess whether or not they have read the background materials. Then, the students watched a video (either dialog or monolog, either individually or collaboratively), along with completing the worksheets assigned for that video. Finally, the students took a post-test.

The videos were assigned to each condition by posting to each group's homepage in Canvas (using the Group function) to prevent students from accessing another type of videos that were not assigned to them. Each video allowed the students to pause, forward, or rewind as needed (Chi et al., 2008; Craig et al., 2004). In addition, the students in each condition were asked to complete the same *Constructive* worksheet while watching the videos, which allowed them to be more generative based on the ICAP framework (Chi, 2009).

The instructors and the tutees

The two instructors, one male biology instructor and one female mathematics instructor, were recruited from the biology and mathematics departments of the same institution. They were selected based on a few criteria, such as their teaching experience, the topics of their courses and course enrollment. Both instructors had more than 10 years of teaching experience and had offered the selected course multiple times. The mathematics instructor served on the course development committee that supervised the content covered in the calculus course of Study 2.

The instructors were also given an orientation on the purpose of this study and the expectation of their commitment to ensure that they were on the same page with the research team. Specifically, the instructors were informed of the various types of videos they would create and the criteria for identifying tutees.

For each of the two courses, two tutee–students, one male and one female, were selected from those who

enrolled in a previous offering of the course by the instructors, to serve as the “tutee” in these tutoring dialog videos. Two criteria were used to select the tutees: (1) having a medium level of performance in their previous courses, and (2) being comfortable with articulating their thought processes.

Materials

The instructional materials used in the two studies were: (1) the pretests and post-tests, (2) the reading materials the instructors assigned, (3) weekly quizzes; (4) six pairs of dialog/monolog videos; and (5) worksheets with problems that resembled those presented in the videos. All materials were created by the course instructors (see examples in Table 6).

Pretests and post-tests

Pretests and post-tests in each study were created/curated by the instructors to measure student learning in respective units. Pretests were designed to assess observing students' prior knowledge. Students received attendance points after completing the pretests. The post-tests included all the pretests items plus additional items addressing a range of difficulty levels. Post-tests were administered as a formal course exam and counted toward their course grade.

Instructors were asked to create both explicit and implicit items for pretests and post-tests. Explicit subsets included questions that required simple factual recall such as finding relevant information directly presented in the videos or combining content from various segments of videos. In contrast, implicit subsets consisted of deep-level questions that students were required to make inferences, undergo complex calculations and apply concepts to solve problems. Once mutual agreement was reached between the researchers and the instructor on a subset of the questions, the instructor continued to label the remaining test items as explicit and implicit questions. Given both instructors' teaching experience, they had no difficulty differentiating implicit from explicit questions.

Reading materials

The biology instructor for Study 1 selected and assigned the reading materials each week. The reading materials basically covered the topics addressed in the videos. Due to the nature of mathematics, instead of assigning reading materials, the instructor for Study 2 offered a short Zoom lecture focused on explaining basic mathematics concepts used in the videos.

Weekly quizzes

The weekly quizzes included multiple-choice questions about the content of the reading materials. Independent

samples *t*-test results revealed that there were no statistically significant differences in students' weekly quiz scores between different conditions in biology and mathematics courses, thus the quiz findings will not be further reported below.

Instructional videos

Before recording any videos, we presented guidelines to the instructors on how to tutor based on the ICAP framework, such as eliciting responses from students rather than providing long explanations, asking questions to students, and providing hints. This guideline was presented via Zoom and took approximately 45 min with about 25 slides. Questions from the instructors on tutoring and the ICAP framework were answered before recording started.

For each study, six pairs of dialog and monolog videos were recorded by the course instructor using Zoom. Each pair of videos covered the same content, consisting of one topic, such as animal behavior for the biology course. The instructors created presentation slides about the content as well as problems to solve in the video. On average, there were four problems (activities or questions) within each video.

To create the dialog videos, the instructor tutored a tutee–student on learning the content of the presentation slides. The instructor worked with one tutee to record the first three sessions and then recorded the remaining three sessions with the second tutee. To avoid tutees getting tired, they were asked to record the three sessions spanning over 3 days.

In solving the problems, both the instructor and the tutee–student used a Wacom Tablet to draw diagrams and demonstrate the steps of calculation. The instructor provided explanations on concepts or questions, gave elaborative and short feedback on tutee–students' answers, and offered question prompts to guide students' thinking processes. Tutees in the dialog videos were free to ask questions and write on the whiteboard in Zoom during the conversations. In short, the tutoring sessions were recorded as dialog videos with the presentation slides and the annotations as Zoom backgrounds.

The same set of slides were used to create the monolog videos. However, the instructor used a lecture-style presentation to teach the STEM concepts and explain how to solve the problems. All instructional videos used in both conditions were unscripted, in the sense that the instructors were asked to cover the content in each session in their normal way of giving a lecture or facilitating a tutoring session, in the way advised by our guidelines.

The length of the two types of videos varied slightly due to the time spent by the instructor on providing feedback to tutees and answering tutees' questions in

the dialog videos. However, we asked the instructors to ensure that both types of videos cover the same key concepts and should fully address each problem.

Worksheets

Worksheets were created by the instructors to engage students in *generative/Constructive* learning in the ICAP sense while watching videos across all conditions. The items in the worksheets were similar but not identical to those discussed in the videos. For example, for the topic of the theory of evolution, a question in the video asked students to explain what it would take for humans to evolve wings, whereas a similar question in the worksheet asked students to explain what it would take to evolve humans without ears. On average, there were four problems (activities or questions) within each of the six worksheets to correspond to the four problems presented in the videos. Instructors in both types of videos prompted the observing students to pause and do the worksheet problems, allowing them sufficient time to respond before resuming the video. This setting was to prevent students from directly copying the solutions and ensure that they practice solving the questions using the concepts learned from the videos.

To ensure that students completed worksheets, all worksheets were included as course assignments which they needed to submit on time. The instructors also sent frequent reminders during each week to reinforce the requirement of completing the worksheets while watching videos. In the end, over 94% of students submitted their worksheets in those two studies.

Study 1: observing dialog and monolog videos in a biology course

This study examined the impact of observing dialog and monolog videos in two classes of an online biology course taught by the same instructor, over two semesters. Within the first semester, students observed individually: half of the students watched the monolog videos (Condition 1) and the other half watched the dialog videos (Condition 2). In the second semester, students observed collaboratively: half watched the monolog videos (Condition 3) and the other half watched the dialog videos (Condition 4). The condition of whether observing individually or collaboratively in the first semester was randomly assigned, followed by observing in an alternative way in the second semester.

Participants

A total of 416 college undergraduates from the two course sections consented to participate in this study. They were mostly sophomore, junior and non-biology

majors. Students who enrolled in this course but did not agree to participate could still access all materials designed for this study, but their data were excluded. Then, within each course section, a stratified random sampling approach was used to assign students to either observing dialog or monolog conditions, resulting in a sample of 114 students in each of Conditions 1 and 2, and 94 students in each of Conditions 3 and 4.

Materials and procedures for study 1

The pretest included 24 multiple-choice questions (twelve explicit and twelve implicit), covering materials taught across the six lessons. The post-test had a total of 31 questions, including all the pretest questions and seven new multiple-choice questions (one explicit and six implicit).

The biology instructor produced two sets of videos focused on animal behavior and four sets of videos centered around population ecology. The same six videos were used across the four conditions. The average length of a dialog video was 36 min, and the average length of a monolog video was 26 min.

For Conditions 1 and 2 in which students watched the videos individually, each student also completed all the six worksheets corresponding to the six videos, with a total of 29 questions. For the two collaboratively observing Conditions 3 and 4, each randomly paired students were asked to schedule two synchronous online meetings each week, during which each pair watched videos and completed the same 29 questions on the six worksheets together on Zoom. There was no restriction on the length of a synchronous meeting, but each pair was required to collaboratively complete each of the six worksheets through discussions. Once a meeting adjourned, each pair submitted the recording of their meetings to Canvas as a required course assignment.

Results

To confirm that the monolog and dialog videos were in fact different, we spot-checked the videos by transcribing and then coding the instructor's instructional moves in two pairs of videos into four categories: explanations, elaborative feedback, short corrective feedback, and scaffolding prompts (Chi et al., 2001, 2008). Table 1 shows each type of moves made by the instructor in the two pairs of dialog and monolog videos. A chi-square test result ($\chi^2=68.17$, $df=3$, $p<0.05$) confirmed that the two types of videos did in fact contain different patterns of instructional moves. The adjusted residual (see Table 1) suggested the differences between the two videos in the number of explanations, elaborative feedback, short feedback, and scaffolding prompts were statistically significant.

Table 1 Frequency of instructional moves by the instructor in two pairs of videos

Videos	Explanation	Elaborative feedback	Short feedback	Scaffolding prompts
Dialog	92 (-8.3)*	30 (5.0)	28 (4.8)	11 (2.9)
Monolog	120 (8.3)	0 (-5.0)	0 (-4.8)	0 (-2.9)

*Adjusted residual (in parenthesis) values ($Z > 1.96$) showed significant differences in two videos

Descriptive statistics including the means and standard deviations for the post-test in the four conditions are provided in Table 2. There were no statistically significant difference in the pretest scores across all four conditions, $F(3,412)=0.87, p=0.46$. The homogeneity assumption was assured by the result of the Levene Test for Equality of Variances for the four conditions, $L(3, 412)=1.84, p=0.139$. A set of ANCOVA, with the pretest score percentage as the covariates were performed.¹ Planned comparisons were performed to examine any significant main effect.

Comparing learning from observing monolog and dialog videos

On average, students learned significantly more from watching dialog than monolog videos ($M=69.07$ vs 65.65 , Table 2), $F(1,411)=11.21, p<0.05, \eta^2=0.03$. This is true whether students watched individually (the difference between Conditions 1 and 2, $F(1,225)=7.05, p<0.05, \eta^2=0.03$) or collaboratively (difference between Conditions 3 and 4, $F(1, 185)=4.44, p<0.05, \eta^2=0.02$).

On average, students also learned significantly more from watching collaboratively than individually ($M=67.88$ vs 66.93 , Table 2), $F(1,411)=4.95, p<0.05, \eta^2=0.01$. However, within each video type, the advantage of watching collaboratively over individually was not significant.

Table 2 Means scores and standard deviations for post-tests of each condition in study 1

		Monolog videos Mean (SD)	Dialog videos Mean (SD)	Average
Individually (n = 114)	Pretest	Condition 1 62.52 (16.60)	Condition 2 61.97 (16.80)	62.25 (16.68)
	Post-test	65.22 (15.83)	68.64 (12.90)	66.93 (14.50)
Collaboratively (n = 94)	Pretest	Condition 3 59.91 (15.61)	Condition 4 59.50 (15.86)	59.71 (15.69)
	Post-test	66.17 (16.42)	69.59 (14.26)	67.88 (15.43)
Average	Pretest	61.34 (16.18)	60.86 (16.40)	
	Post-test	65.65 (16.06)	69.07 (13.50)	

¹ The statistical tests were performed using SPSS Version 27.

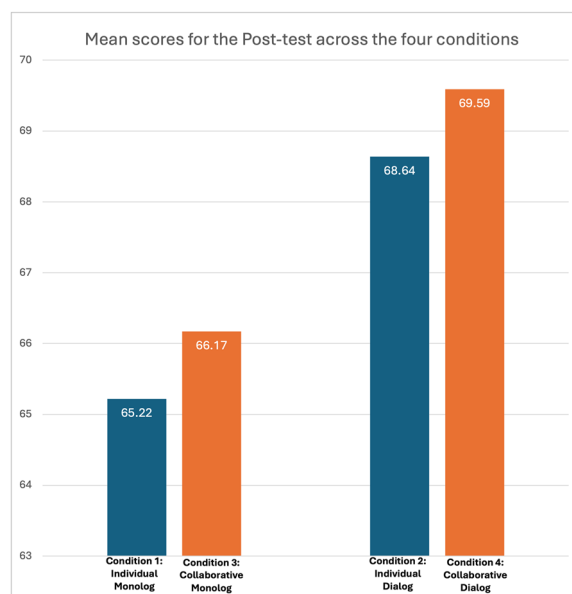


Fig. 2 Mean scores for the post-test across the four conditions

Figure 2 shows the mean percentage of the post-test scores across the four conditions, and an ANCOVA reported a significant difference, $F(3, 411)=5.43, p<0.05, \eta^2=0.04$. The post-hoc comparison results indicated that students collaboratively observing dialog videos ($M=69.59$) earned a significantly higher post-test percentage score than those who observed monolog videos, either collaboratively ($M=66.17$) or individually ($M=65.52$). No other post-hoc comparisons were significant.

Comparing student performance in various levels of questions

Table 3 shows the post-test scores for explicit and implicit questions for each of the four conditions. For explicit

Table 3 Descriptive statistics for explicit and implicit questions in the post-tests across conditions

Post-test questions		Monolog videos Mean (SD)	Dialog videos Mean (SD)	Average
Individually (n = 114)	Explicit	9.54 (2.27)	9.96 (1.72)	9.75 (2.02)
	Implicit	7.66 (2.26)	7.92 (2.02)	7.79 (2.14)
Collaboratively (n = 94)	Explicit	9.36 (2.17)	9.82 (1.89)	9.59 (2.04)
	Implicit	11.15 (3.33)	11.74 (3.12)	11.45 (3.23)
Average	Explicit	9.46 (2.22)	9.89 (1.79)	
	Implicit	9.24 (3.28)	9.65 (3.20)	

questions, overall, students performed significantly better when they observed the dialog videos than the monolog videos ($M=9.89$ vs $M=9.46$, Table 3), $F(1,411)=8.46$, $p<0.05$, $\eta^2=0.02$. The significant difference in the average performance on the explicit questions was obtained whether students watched individually (the difference between Conditions 1 and 2, $F(1, 225)=4.50$, $p<0.05$, $\eta^2=0.02$) or collaboratively (the difference between Conditions 3 and 4, $F(1, 185)=3.96$, $p<0.05$, $\eta^2=0.02$).

However, the difference in the average performance on the explicit questions between students observing videos collaboratively and individually (9.59 vs 9.75) was not significant. Similarly, there was no significant difference in explicit question scores when observing the same type of videos, either dialog (9.82 vs 9.96) or monolog videos (9.36 vs 9.54).

For implicit questions, overall, students observing videos collaboratively earned a significantly higher score in answering implicit questions than those observing videos individually (11.45 vs 7.79), $F(1, 411)=193.47$, $p<0.05$, $\eta^2=0.32$). The advantage of observing collaboratively in answering implicit questions carried over when students observing the same type of videos, either dialog (11.74 vs 7.92), $F(1, 205)=167.16$, $p<0.05$, $\eta^2=0.45$) or monolog (11.15 vs 7.66), $F(1, 205)=151.84$, $p<0.05$, $\eta^2=0.43$) videos.

On average, for implicit questions, students observing dialog videos performed significantly better than those observing monolog videos (9.65 vs 9.24, see Table 3), $F(1,411)=4.72$, $p<0.05$, $\eta^2=0.01$. The significant difference occurred between students collaboratively observing dialog videos ($M=11.74$, Condition 4) and those individually observing monolog videos ($M=7.66$, Condition 1).

Study 2: observing individually in a mathematics course

Study 2 sought to replicate more specifically the finding in Study 1 that watching dialog videos individually was more beneficial than watching monolog videos

individually in another STEM content domain, mathematics. Replicating this finding across various STEM domains validates observing dialog videos individually as an innovative, easy-to-implement instructional format.

Methods

A total of 89 students enrolled in the online Calculus course consented to participate in this study, with 45 students assigned to the monolog group and 44 to the dialog group. The data from 15 students who did not complete pretests or post-tests were removed for analysis, ending up with a sample ($N=74$) of 36 students observing monolog videos and 38 students observing dialog videos. A majority of participant students were freshman from business majors.

Materials and procedures for study 2

The mathematics instructor and two tutees who took the calculus course from the previous semester developed six sets of videos. The average lengths of a dialog video and a monolog video were 22 min and 15 min, respectively, covering the topics of Definite Integral, Substitution, Numerical and Graphical Viewpoints. The dialog/monolog videos only focused on demonstrating the calculus problem-solving process. Different from Study 1, instead of assigning reading materials, the mathematics instructor delivered a short live lecture via Zoom before students watched the assigned videos. The lecture presented fundamental definitions and formulas as the basis for solving calculus problems. Therefore, the quizzes administered after the Zoom lectures were about the mathematical concepts presented in the Zoom lectures. While watching the dialog or monolog videos, students were required to solve the problems on worksheets that were similar but different (such as using different values or variables) calculus problems. On average, there were four problems on each worksheet.

Study 2 lasted 2 weeks due to logistic requirements, leaving the students attending three online sessions each week. Given the time restriction on mathematics tests imposed by the department, the pretests/post-tests were relatively short. The pretest contained 10 questions (two explicit questions and eight implicit questions), and the post-test had 13 questions (adding three new implicit questions). Explicit questions asked students to recall information from the video and use the same formula to answer the question, while implicit questions expected students to make inferences based on the content provided in the video and then use the inference to complete complicated calculation. The instructor curated the test items from a test item bank and categorized the questions as explicit or implicit, with guidance from the researchers.

Results

Table 4 shows the mean scores percentages and standard deviations of the pretest and the post-test from the two groups. The independent samples *t*-test results showed no statistically significant difference in pretest scores between the two conditions ($p=0.47$) and the homogeneity assumption was met, $L(2, 72)=1.09$, $p=0.31$. ANCOVA, with the pretest score percentage as the covariate, showed that the students who individually watched dialog videos gained a significantly higher post-test score percentage than those who individually watched monolog videos (57.80 for dialog vs 46.16 for monolog), with a medium effect (Cohen, 1988), $F(1, 71)=8.64$, $p<0.05$, $\eta^2=0.11$. Table 5 shows students

observing dialog videos performed significantly better in implicit questions (5.76 for dialog vs 4.42 for monolog) when the pretest score percentages were controlled, $F(1,71)=8.24$, $p<0.05$, $\eta^2=0.10$, with a small effect. The differences in the post-test scores percentages on explicit questions between the two conditions did not reach the level of significance.

Discussion

The goal of the studies described in this paper was to see if we can replicate the previous laboratory findings in authentic online learning context about a new instructional resource, tutorial dialog videos, because we have previously shown that students learn more from watching dialog videos than lecture-style monolog videos. To review, there are four important findings in our current studies that reinforce the suggestion that viewing tutorial dialog videos may be a good instructional resource, especially for online learning. First, the results reported here replicated the advantage of observing tutorial dialog videos over lecture-style monolog videos from a laboratory context (Chi et al., 2008) in authentic STEM classroom context, across two different content domains (e.g., biology and mathematics). Second, in particular, the learning advantage occurs for the more difficult implicit questions in both Studies 1 and 2, suggesting deeper understanding. Third, watching collaboratively had a learning advantage over watching individually, again replicating our prior findings. However, the advantage of watching dialog videos (over monolog videos) was greater than the advantage of watching collaboratively (vs individually). Given

Table 4 ANCOVA results and descriptive statistics for each condition in study 2

	Monolog		Dialog		F	p	df	η^2
	M	SD	M	SD				
Pretest	10.97	15.80	8.55	12.57	8.64	0.00*	1.71	0.11
Post-test	46.16	16.90	57.80	18.94				

Table 5 Number of different levels of questions answered correctly in each condition for study 2

Difficulty	Monolog		Dialog		F	p	η^2
	M	SD	M	SD			
Explicit					1.66	0.20	0.02
Pretest (n=2)	0.42	0.60	0.26	0.45			
Post-test (n=2)	1.53	0.61	1.68	0.47			
Implicit					8.24	0.005*	0.10
Pretest (n=8)	0.68	1.17	0.59	1.14			
Post-test (n=8+3)	4.42	1.90	5.76	2.32			

*Significance at the 0.05 level

that learning collaboratively has been shown repeatedly in the literature to be a beneficial mode of learning, the greater benefit of watching dialog videos further commends the utility of dialog videos for instruction. Fourth, the finding that college students watching dialog videos individually learned better than watching monolog videos individually, suggests that watching dialog videos individually, can be implemented relatively easily in an online context.

What factors could mediate the benefits of observing dialog videos? Several factors were considered in Chi et al. (2017). The dialog videos are usually longer, suggesting that the additional time that the tutors may have taken tailoring their explanations to tutees' individualized needs in the tutoring sessions, may be the source of students' improved learning. However, Chi et al. (2017) found no significant correlation between the frequency of the tutor's move (e.g., providing feedback or scaffolds) and student learning. Rather than the length or frequency of explanations in each video, we suggest that students who observed dialog videos learn better because of the following five reasons, based on a synthesis of our findings and other empirical evidence.

First, observing students are more likely to understand statements or remarks made by tutees in the dialog video given that they may have a similar level of relevant expertise, which Chi (2013) referred to as a representational match. To substantiate this point, Chi (2013) examined students' referral statements to the instructor's utterances in monolog and dialog videos and found different patterns between those two groups. Specifically, students observing dialog videos tended to elaborate on what the tutors and the tutees both discussed, whereas it is only possible to comment on what the tutor said, in the monolog videos. Having the tutees as the target of referrals in the dialog videos may be advantageous as students and tutees may share similar level of understanding; that is, they are more likely to have a representational match, than between a student and the tutor. Overall, the zone of proximal representational match may allow observing students to understand the tutees better than the tutors, and thus learn more from the tutees in the dialog videos than the instructors in the monolog videos.

Second, students observing dialog videos have opportunities to imitate tutees' manner of performing constructive learning, such as asking questions and generating substantive comments (Chi, 2013; Chi et al., 2008; Rummel & Spada, 2005). Observing dialog videos may empower them with constructive learning skills such as asking their own questions, seeking substantive answers, and elaborating on their understanding.

Third, research has indicated that watching the tutee's struggles or committing errors in the dialog videos can

help facilitate learning, particularly by encouraging observing students to be more *Constructive* and *Interactive* in the ICAP sense (Chi & Wylie, 2014). For example, Schunk et al. (1987) found that children learned better when watching a video of subtraction that featured a struggling student's problem-solving procedures than a competent peer solving the same problem. Similarly, Muller et al. (2008) suggested that the tutor-tutee conversations in the dialog videos may reveal observing students' challenges in understanding the topic if the tutee discussed similar challenges. However, Chi et al. (2017) did not find a correlation between tutees' faulty claims and observers' learning gains, so the plausible advantage of watching a tutee struggle might be mediated by self-efficacy (Shunk & Zimmerman, 2007) rather than the advantage of overhearing tutees' misunderstanding. That is, Ormrod (2020) suggested that by allowing students to watch a "model" who had difficulty at the beginning but then managed to solve a problem in the end, may build the observing students' self-efficacy, thus promoting their learning.

Fourth, Chi et al. (2017) suggested that students observing dialog videos may become more intrinsically motivated to invest effort in the pursuit of a correct answer if they identified the errors in the tutee's talks. That is, they found that the tutee's struggles or errors in the dialog videos appear to encourage the observing students to be more *Constructive* and *Interactive* while trying to solve a problem and thereby becoming more elaborative and knowledgeable in the topic discussed in the dialog videos.

Fifth, most importantly, the ICAP theory suggests that tutoring is generally an effective form of instruction because the context allows students greater opportunity to be generative/*Constructive*. This interpretation of the benefit of tutoring was confirmed by our prior study in which normal tutoring was compared with a form of restricted tutoring in which tutors were forbidden from giving scaffoldings and explanations; instead, the tutors were only permitted to prompt students for responses. This led to equivalent learning outcomes among the tutees in both conditions, suggesting that the advantage of tutoring rests on the greater opportunities for tutees to be generative (Chi et al., 2001). It is not far-fetched to suggest that perhaps the observing students will also be more generative, upon hearing either the tutor's comments or the tutee's comments. In fact, the analyses of collaborative observing students' conversations do show that they tend to be more generative and collaborative (Chi et al., 2017). In sum, the preceding five explanations suggest plausible reasons for why observing dialog videos might enhance learning more so than observing monolog videos.

The findings of this study have significant implications for designing effective instructional practices in STEM education. One notable implication is that, watching dialog videos, which are unscripted, can be implemented as an easy and effective way of teaching STEM subject content online. Without the need to write scripts, instructors can easily record tutoring conversations with students in a natural setting. Second, dialog videos can also be effective for students learning individually compared to didactic monolog videos. Third, it is also easy to find students in one's class to participate as the tutee in the dialog videos as students know they can benefit from being tutored.

Conclusions

Overall, our findings provide opportunities for scaling-up dialog videos as a low-cost effective instructional format that may have a high impact on student learning and can be easily implemented at scale. Instructors may integrate dialog videos in online courses easily to facilitate student learning.

Appendix

See Table 6.

Table 6 Sample questions

Question type	Sample question
Pretest and post-test	
Implicit	You are working with local conservation agencies to monitor a key population of Gila monsters living in the Sonoran Desert (pictured right). Previous studies estimate that the typical carrying capacity of Gila monster populations is 50 individuals. You estimate the population size of Gila monsters in this area to be 25 individuals in 2009. Based on your research, you discover that Gila monster populations typically have a per capita growth rate of 0.25. Based on this information, estimate the number of Gila monsters in the population in the year 2010. If your calculation has decimals, round up if your decimal is 0.5 or larger, round down if your decimal is below 0.5. Select only <i>ONE</i> answer choice A. 2010 = 26 B. 2010 = 28 C. 2010 = 31 D. The answer is not listed in answers A–C
Implicit	A pond ecosystem in an open field begins to be shaded by the growth of trees around its perimeter. Predict changes in this pond after the trees grow large enough to completely shade the pond. Select only <i>ONE</i> answer choice The population sizes of all organisms within the pond food web will decrease A. The population sizes of all organisms within the pond food web will increase B. The diversity of organisms within the pond food web will increase C. The pond food web will not be affected by shading
Explicit	If a bird song promotes an individual's ability to attract a mate and reproduce, then the song has a(n) _____ function A. Causative B. Adaptive C. Mechanistic D. Assumptive E. Historical
Weekly quiz	
Multiple-choice question	A severe drought is an example of a density-_____ factor that can influence population size. The availability of nest sites is an example of a density-_____ factor A. dependent; independent B. independent; dependent C. dependent; dependent D. independent; independent
Worksheet	
Multiple-choice question	Why do prairie dog mothers ruthlessly kill ground squirrel pups that turn up near their burrows? A. The ground squirrels provide supplementary protein B. The ground squirrels share a similar ecological niche and compete for grass and plant food with the babies of the mother prairie dog C. The ground squirrels spread disease that could threaten her babies D. The ground squirrels attract additional predators that could harm the prairie dogs E. Any of the above could be correct. We need more data...
Open-ended question	Can you think of any other reasons besides kin selection or group selection that might explain cooperation among animals?

Acknowledgements

This work was supported by the National Science Foundation grant #1915150. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency. We would like to extend our deepest thanks to the instructors and students who participated in this study. We also sincerely appreciate the constructive comments from all the editors and reviewers. The authors acknowledge our funding agency, National Science Foundation. The authors also express our sincere gratitude to the instructors and participating students who contributed to our project.

Author contributions

YQ: conceptualization, literature review, methodology, data analysis, writing—original draft, and writing—review and editing. YH: conceptualization, literature review, methodology, and writing—review and editing. MC: conceptualization, literature review, methodology, and writing—review and editing.

Funding

This work was supported by National Science Foundation under NSF IUSE Award #1915150.

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board at Arizona State University. All participants provided informed consent prior to their participation in this study.

Consent for publication

We hereby provide consent for the publication of the manuscript detailed above, including any accompanying images or data contained within the manuscript that may directly or indirectly disclose our identities.

Competing interests

The authors have no competing interests to declare.

Received: 9 July 2023 Accepted: 30 August 2024

Published online: 28 September 2024

References

- Auerbach, A. J. J., & Andrews, T. C. (2018). Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: A large-scale qualitative investigation of instructor thinking. *International Journal of STEM Education*, 5(1), 19. <https://doi.org/10.1186/s40594-018-0112-9>
- Barbieri, C., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences*, 48, 36–44.
- Barlow, A., & Brown, S. (2020). Correlations between modes of student cognitive engagement and instructional practices in undergraduate STEM courses. *International Journal of STEM Education*, 7(1), 18. <https://doi.org/10.1186/s40594-020-00214-7>
- Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Example problems that improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, 25, 24–34.
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H. (2013). Learning from observing an expert's demonstration, explanations and dialogues. In J. J. Staszewski (Ed.), *Expertise and skill acquisition: The impact of William G. Chase* (pp. 1–28). Psychology Press.
- Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating a theory of cognitive engagement into practice. *Cognitive Science*, 42(6), 1777–1832. <https://doi.org/10.1111/cogs.12626>
- Chi, M. T. H., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences*, 26(1), 10–50. <https://doi.org/10.1080/10508406.2016.1204546>
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutoring collaboratively: Insights about tutoring effectiveness from vicarious learning. *Cognitive Science*, 32, 301–341. <https://doi.org/10.1080/03640210701863396>
- Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49, 1–25. <https://doi.org/10.1080/00461520.2014.965823>
- Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, 12(4), 425–434.
- Craig, S., Driscoll, D., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia*, 13, 163–183.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson.
- Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, 12, 45–58.
- Lobato, J., Gruver, J., & Foster, M. (2023). Students' development of mathematical meanings while participating vicariously in conversations between other students in instructional videos. *The Journal of Mathematical Behavior*, 71, 101068. <https://doi.org/10.1016/j.jmathb.2023.101068>
- Makwana, D., Engineer, P., Dabhi, A., & Chudasama, H. (2023). Sampling methods in research: A review. *International Journal of Trend in Scientific Research and Development*, 7, 762–768.
- Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction*, 13(3), 315–372.
- Muldner, K., Lam, R., & Chi, M. T. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, 106(1), 69–85.
- Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, 24(2), 144–155.
- National Center for Education Statistics. (2022). *Distance Learning: Fast Facts*. <https://nces.ed.gov/fastfacts/display.asp?id=80>
- Ormrod, J. E. (2020). *Human learning: Principles, theories, and educational applications* (8th ed.). Merrill Publishing Co.
- Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534–574.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences*, 14(2), 201–241.
- Schunk, D. H., Hanson, R. A., & Cox, P. D. (1987). Peer-model attributes and children's achievement. *Journal of Educational Psychology*, 79, 54–61.
- Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, 23(1), 7–25. <https://doi.org/10.1080/10573560600837578>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Weinberg, A., Corey, D. L., Tallman, M., Jones, S. R., & Martin, J. (2022). Observing intellectual need and its relationship with undergraduate students' learning of calculus. *International Journal of Research in Undergraduate Mathematics Education*. <https://doi.org/10.1007/s40753-022-00192-x>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yingxiao Qian is a Clinical Assistant Professor in the Department of Leadership, Learning Design, and Inquiry at the University of South Carolina. Her research interests include designing technology-enhanced learning environments that support students' real-world problem solving, computational thinking, and active learning in STEM education.

Yi-Chun Hong is with the Mary Lou Fulton Teachers College at Arizona State University. Her work focuses on designing effective, meaningful learning environments that promote active learning and develop students' ill-structured problem-solving abilities.

Micheline Chi is a Regents Professor and the Dorothy Bray Endowed Professor of Science and Teaching in the Mary Lou Fulton Teachers College at Arizona State University. Her research interests focus on active learning and instructional videos for online learning. She has developed a framework for active learning called ICAP.