

Learning from error episodes in dialogue-videos: The influence of prior knowledge

Lu Ding

Eastern Illinois University

Katelyn M. Cooper, Michelle D. Stephens, Michelene T.H. Chi, Sara E. Brownell

Arizona State University

In laboratory study environments, dialogue-videos, or videos of a tutor and a tutee solving problems together, have been shown to more effectively improve student learning than monologue-videos, or videos of tutors solving problems alone. Yet, few studies have replicated these findings in the context of authentic university classrooms. Here, we investigate the impact of dialogue-videos, and more specifically the effect of errors made by tutees in dialogue-videos, on student learning in the context of an undergraduate biology course. To understand why, we investigated students' effort spent on watching videos, perceived influence of dialogue-videos, and worksheet completion rates. We found that higher-performing students perceived that they used the dialogue-videos to review content. We also found that higher-performing, but not lower-performing, students learned better from dialogue videos where tutees made errors. We also discuss the complexities of replicating laboratory studies in the classroom and implications of our findings.

Implications for practice or policy:

- Tutee errors can be intentionally included in dialogue-videos to promote student learning.
- When students lack the necessary prior knowledge, monologue-videos may be more effective in presenting the course content.
- When using dialogue-videos, instructors can encourage students to collaborate to resolve any confusion in time to maximise the benefit of dialogue-videos in teaching and learning.

Key words: dialogue-videos; monologue-videos; errors; prior knowledge; mixed methods research

Introduction

Videos are a predominant content delivery format for fully online, blended, and flipped classrooms (Scagnoli et al., 2017). Undergraduates have reported that they enjoy learning from videos (Evans & Cordova, 2015), and prefer videos to readings as a form of content delivery (Cooper et al., 2018). In contrast to face-to-face lecturing, videos often have the advantage of providing students with the opportunity to pause, rewind, and skip ahead to direct their learning experience (Fyfield et al., 2019). However, a pervading challenge is that videos are often didactic lectures (Fyfield et al., 2019). That is, students listen passively to instructors; however, studies have shown that didactic lecturing is less effective than student-centered active learning (Freeman et al., 2014) and that students can learn more from answering questions than passive absorption of knowledge.

Tutoring is regarded as one of the most effective forms of student-centered learning because of the profound effect it has on student learning (Bloom, 1984; VanLehn, 2011; Wood & Tanner, 2012). Specifically, according to Chi et al. (2001), tutoring scaffolds a student's understanding of a specific concept or problem by allowing the student to actively engage in the learning process and construct their own knowledge; the tutor asks questions and provides personalised feedback until the student fully understands the content they are learning. In fact, student learning from tutoring has been largely attributed to the dialogues between a tutor and tutee, which include asking and answering questions, as well as the guidance that the tutor provides (Chi et al., 2001).

Although tutoring is considered a gold standard of instruction (Bloom, 1984), one-on-one tutoring is not cost effective and has been presumed to be an impossible way of teaching content to a large cohort of students in a university classroom (Van der Kleij et al., 2015). However, one solution to this problem is to

create videos of a tutor tutoring a tutee in an effort to enhance the learning of students who watch the videos. Even though the observing student would not have personalised feedback from the tutor, they could observe the tutee asking questions and getting their questions answered by the tutor. For consistency, hereafter we call any video of dialogues between tutors and tutees dialogue-videos, we call videos of only a tutor presenting a concept or demonstrating how to solve problems monologue-videos, and we use the term observing students to refer to students who watch the videos.

Previous studies on dialogue-videos

Previous studies have demonstrated that students can learn equally well from face-to-face tutoring and dialogue-videos. Specifically, one study conducted with undergraduate students found that observing students who watched dialogue-videos of tutors guiding tutees to solve physics problems in dyads (i.e., two observing students watched the dialogue-videos and solved the same problems presented by the tutors in the video together) learned as well as the tutees who were tutored in the videos (Chi et al., 2008). A similar result was reported in Muldner et al.'s (2014) study which focused on chemistry questions. To assess the importance of a tutee being present in the videos, they further compared the student performance on the pre- and post-tests of observing students in dyads watching dialogue-videos with the performance on the tests of observing students in dyads watching monologue-videos. The authors found that the observing students learned significantly better when they watched dialogue-videos than monologue-videos. These findings beg the question of why students learned better when watching dialogue-videos. Chi and colleagues (2017) conducted a further analysis on Muldner et al.'s study, and revealed that the conflict episodes, or instances where a tutee expressed a misconception that was followed up by a tutor's correction of the idea, generated more constructive comments between the pairs of students watching videos (dyads) and resulted in higher performance on the post-tests.

Several laboratory studies have also provided some evidence that observing students watching dialogue-videos alone may still positively affect student learning. For example, Driscoll et al. (2004) found that students who individually listened to the dialogues between a virtual tutor and a tutee discussing computer literacy topics wrote significantly more content in their essays than students who listened to only a virtual tutor presenting the topics. Additionally, Muller et al. (2008, 2007) found that observing students who individually watched dialogue-videos of a tutor and a tutee discussing the common misconceptions about Newton's First and Second Laws outperformed the students who individually watched monologue-videos where only the tutor explained the correct facts without refutations on the tests.

Learning from cognitive disequilibrium

Confusion has been shown to be one of the most frequently occurring emotions in tutoring sessions when students ask and answer questions of the tutor (Lehman et al., 2010). Confusion is triggered when students experience cognitive disequilibrium, such as when they reach an impasse, experience discrepancies between their existing knowledge and the new information, or receive contradictory information from different sources during learning, leaving them uncertain about how to proceed (Arguel et al., 2017; Baker et al., 2010; Lehman et al., 2012). Craig et al. (2004) coded students' five affective states (i.e., frustration, confusion, boredom, flow, and eureka) while interacting with an intelligent tutoring system, where a virtual tutor assisted the students answering questions about computer literacy. They found that students who experienced confusion improved significantly more from pre- to post-tests than students who did not. It is hypothesised that deep comprehension occurs when students experience confusion due to the effort that is more likely to be executed to restore the cognitive equilibrium. Similar findings have been reported in replication studies conducted with the same intelligent tutoring system (D'Mello & Graesser, 2011; Graesser et al., 2007).

Why does confusion positively influence student learning? Students do not learn from confusion itself, but rather confusion encourages students to engage in deep learning activities such as reflection, deliberation, and deciding "which opinion had more scientific merit" (D'Mello et al., 2014, p. 155). Confusion also causes students to seek help when a cognitive disequilibrium is encountered and cannot be resolved by oneself (Ryan et al., 2005). Such effortful cognitive activities can result in greater learning. Nevertheless, it should be emphasised that there is no guarantee for this learning to occur. Learning happens only when students can successfully regulate the confusion, and resolve the cognitive disequilibrium (D'Mello & Graesser, 2012). Otherwise, confusion would inflict damage on the process of students newly learning

scientific concepts or could result in students having a false sense of knowing, and thus limiting further mental effort in learning (Muller & Sharma, 2007).

Requisite prior knowledge is required in order for cognitive disequilibrium to lead to deep learning (D'Mello et al., 2014; Zohar & Aharon-Kravetsky, 2005). An event that is beyond a student's zone of proximal development can create hopeless confusion or the potential learning opportunity from confusion can be ignored, which is typically detrimental to learning (Arguel & Lane, 2015). For example, Zohar and Aharon-Kravetsky (2005) compared a teaching method which purposefully introduced cognitive confliction with teaching that did not intentionally introduce cognitive conflict in a face-to-face classroom setting. They found that higher-performing students learned better from the cognitive confliction method, whereas lower-performing students benefited more from direct teaching. They argued that higher-performing students had sufficient prior knowledge and reasoning abilities to recognise and resolve the conflict, but lower-performing students did not have the aptitudes, and thus the lower-performing students could not benefit from the confusion.

However, it is important to note that prolonged confusion can also be detrimental to learning (D'Mello et al., 2014). When students stay confused and without external supports or scaffolding, this persistent confusion can lead to anxiety and possibly despair (Cooper et al., 2018; Zeidner, 2007). Studies have shown that eventually students experiencing persistent confusion will be at risk of disengaging from learning and develop negative emotions such as frustration or boredom (D'Mello & Graesser, 2012; Pekrun et al., 2010). Instead of engaging in cognitive activities, disengaged students are likely to exhibit shortcut learning behaviors such as guessing or looking for direct solutions (Aleven et al., 2006; Baker et al., 2004), and hence in this case, confusion becomes detrimental to learning. Thus, there is both a positive and a negative side to confusion.

The motivation of the current study

While the literature suggests that students benefit from: (a) watching dialogue-videos in dyads in a controlled laboratory setting; (b) watching dialogue-videos in dyads in the context of a course; and (c) watching dialogue-videos individually in a controlled laboratory setting, to our knowledge no studies have explored the effect of students watching dialogue-videos individually in the context of a university course. Laboratory studies are often carried out during a strictly controlled environment, whereas participants in the studies conducted in the context of a classroom often have more leeway. Therefore, there is a need to examine whether laboratory-based cognitive science studies can be replicated in a classroom environment (McDaniel et al., 2017; Mestre et al., 2018). Furthermore, although the aforementioned studies have demonstrated the effectiveness of dialogue-videos on student learning, all required students to view the dialogue-videos in dyads. However, it is often a challenge for students to watch a video together for an assignment online due to internet bandwidth issues and appropriate platforms that allow two students to watch videos simultaneously.

To fill this gap in the literature, we conducted a study to investigate to what extent dialogue-videos support individual observing students' learning in the context of an undergraduate senior level physiology course. In contrast to what had previously been found, we demonstrated that the observing students learned equally well from watching dialogue- and monologue-videos individually (Cooper et al., 2018; Ding et al., 2018). In addition, the majority of the observing students preferred watching monologue-videos compared to the dialogue-videos. The main reason that observing students preferred the monologue-videos is that the errors made by the tutees in the dialogue-videos were confusing, which they perceived inhibited their learning. Yet, we did not explore the effect of these errors on students' understanding of the concepts covered in the course in this previous study.

In this study, we are particularly interested in assessing the impact of errors made by tutees on observing student learning. Considering that prior knowledge plays a critical role when confusion occurs, we also wanted to test to what extent student prior performance determines their understanding of physiology concepts from watching the error episodes presented in dialogue-videos compared to monologue-videos containing the same content, without error episodes. In addition, we further investigated the possible factors that may have contributed to student learning from error episodes. The remainder of the paper is organised as follows. We first provide some details of the research design, and then we report our analyses and results as two parts. In the first part, we report student test results from the error episodes in dialogue-videos.

Particularly, a comparison between student performance on the test questions targeted on the error episodes and the corresponding episodes in the monologue-videos. In the second part, we report the results from the survey data: what factors that may have influenced their learning. Finally, we discuss some implications.

Research design

This study was conducted in an undergraduate level physiology course with 280 students who met in person three times per week (Tuesday, Thursday, and Friday) in 2017 from August to December at a large southwestern research university in the United States. Approval from the ethical committee of the university was obtained before the study started. The instructor made two types of videos covering physiology content for this study: monologue-videos and dialogue-videos. More specific information is outlined in each section below.

Procedure

The study was conducted over 8 consecutive weeks. A counter balancing design was applied, and students were randomly assigned to group A or group B (Shadish et al., 2002). Each group watched one type of video for the first 4 weeks (e.g., monologue) and the other type of video for the second 4 weeks (e.g., dialogue). The videos covered the same content (Table 1). Students watched a video and completed a corresponding worksheet each week for their homework outside of class. The video and worksheet were available to students after class on Tuesday, and students were asked to watch the video after class on Thursday so that they would have 3 days to complete this activity. The worksheets were collected at the beginning of class on Friday and graded for completion. Students then completed an in-class quiz each week by using an online platform during class on Friday on the content presented in the video. After weeks four and eight, students were surveyed about their experience watching the videos.

Table 1

Topics covered in the videos, the group assignment, and the run times of each video (minutes:seconds)

Week and Topic	Group A	Group B
1. Homeostasis	Monologue (20:12)	Dialogue (22:26)
2. Information flow	Monologue (20:47)	Dialogue (25:13)
3. Temperature sensation	Monologue (13:49)	Dialogue (17:15)
4. Action potentials	Monologue (16:48)	Dialogue (19:42)
5. Synaptic cleft	Dialogue (27:15)	Monologue (21:26)
6. Signal transduction	Dialogue (25:18)	Monologue (14:21)
7. Experimental design	Dialogue (16:02)	Monologue (12:13)
8. Leptin signaling	Dialogue (18:18)	Monologue (15:44)

Participants

Among the 280 ($n = 280$) students enrolled in the course, 217 ($n = 217$) consented to participate in the study. Of these, 114 were randomly assigned to group A, and 103 were randomly assigned to group B. The slight difference in numbers between the two groups was due to student withdrawals after the initial groups were assigned at the beginning of the course. No differences regarding student demographics, including gender, race/ethnicity and prior GPA, were detected between the two groups (Table 2).

Table 2

Demographic information of students assigned to the two groups

		Group A (n = 114)	Group B (n = 103)
^a Gender	Male	34.2%	33.0%
	Female	55.3%	57.3%
Race/ethnicity	^b Underrepresented minority	25.0%	22.1%
	^c Non-underrepresented minority	75.0%	77.9%
Average GPA		3.43	3.34

Note. ^a Some students either wished to not reveal their gender or identified as other. ^b Underrepresented minority includes African American, American Indian, and Latino/a students. ^c Non-underrepresented minority includes white and Asian students.

Development of instructional materials and assessments

Video creation

A set of two videos were made for each of the 8 weeks, one monologue-video and one dialogue-video covering identical physiology content. Each video contained two to six physiology worked examples. A worked example consists of a problem, the steps taken to reach a solution, and the final solution. In the dialogue-videos, the instructor tutored a student tutee working through the problems. The instructor first let the tutee attempt the problem, and then corrected any errors made by the tutee or asked questions to guide the tutee to correct an error until a final correct solution was reached. Four students who completed the course in the previous year were recruited to be tutees in the dialogue-videos. The tutees did not review the content covered in the videos beforehand, rather, they were broadly familiar with the topics presented and reacted authentically when presented with the physiology problems. Each tutee filmed two videos: one in the first set of four videos and one in the second set of four. The monologue-videos only presented the instructor alone solving the same set of physiology problems. The length of dialogue-videos varied from 16 to 27 minutes and averaged 21 minutes, and monologue-videos ranged from 12 to 21 minutes with an average of 17 minutes (please see Table 1 for runtimes of each video). The videos were uploaded to the university's learning management system where students can play, pause, and rewind videos; however, students could not fast-forward. Table 3 presents the content covered in a typical episode of a monologue- and the corresponding dialogue-video, and Figure 1 is a screenshot of the videos.

Table 3

A sample episode from a monologue-video and the corresponding dialogue-video

Content presented in monologue-videos	Content presented in dialogue-videos
Instructor: The things we want to be thinking about here are, we're thinking about charges. We're thinking about positive and negative charges and how that's going to impact, the overall, charge of the cell.	Instructor: So, do you want to kind of draw out what you're thinking about?
Instructor: So, will the neuron hyperpolarise or depolarise when sodium enters the cell?	Tutee: Yes, sure. So, if we have a cell and we have sodium in there we have negative charges inside, positive outside. So that's causing the negative resting membrane potential and through a channel. If you have sodium come in you have positive charges coming inside making this less negative inside, and balancing out this charge difference between the interior and outside of the membrane, and that's going to bring it towards zero, and depending on how much sodium comes in it's going to change the positive charge which then it might become positive or maybe close to zero.
Instructor: When sodium enters the cell, we know that we have, so if we have our kind of generic picture of our neuron, cell membrane, we're going to have a lot of sodium outside the cell, a lot of potassium inside the cell, and these concentration gradients are set up using ATP and other processes.	Instructor: Mhmm, so um why zero?
Instructor: In a normal action potential, when voltage-gated sodium channels open, we're going to have a lot of sodium that comes in and that's what leads to that rising phase of the action potential. So, this is going to cause a depolarisation event to actually occur. So, in terms of thinking about the, the membrane potential, it's going to get more positive. So, because these positive ions are actually coming in, right? So that's what's happening there.	Tutee: Well right now it's, if we had zero here, it's down here at negative 68, if we become more positive, you're going to go up towards zero, but I mean its variable depends upon how much sodium comes in it could go above or below or at zero.
	Instructor: Okay. Do you have a sense for, so, you mentioned that it goes in through a channel? What channel does it go in through?
	Tutee: It'd be in through. Well I mean, it depends, I mean you could have it go through a leak channel if you add a bunch of sodium outside. But it you could have it go through a

sodium voltage-gated channel as well. If you depolarise and somehow open those channels

Instructor: Mhmm, so basically, it's gonna be a sodium specific channel though, right?

Tutee: Right

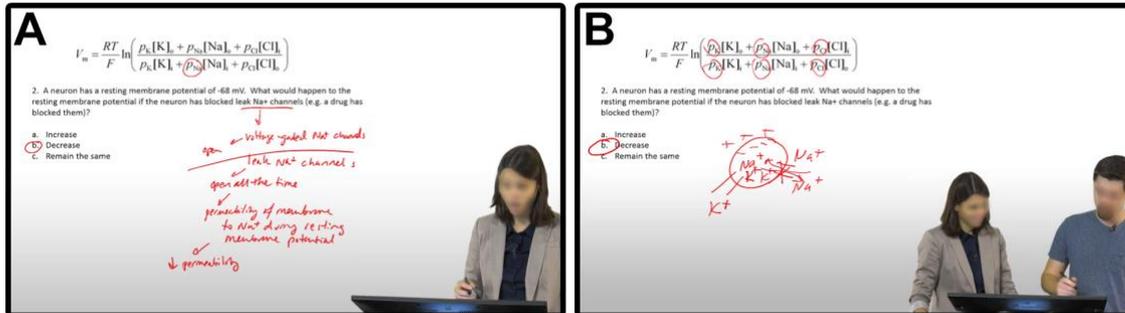


Figure 1. Screenshots of a monologue-video and a dialogue-video on the same topic

Worksheet creation

To maximise student learning by encouraging student engagement while watching the videos (Freeman et al., 2014), a worksheet containing the same questions presented in the video was created for students to complete. Therefore, the student had to engage with the video in order to know how to answer the question.

Quiz creation

Students' learning from the videos was measured by eight quizzes created by the instructor. Students completed a quiz each week after watching each video. Each quiz consisted of 10 to 12 multiple-choice questions, totaling 89 questions. The quizzes were first piloted with 168 students in the same physiology course offered in the year before the present study, and we observed a ceiling effect in student scores on the quizzes. We revised the quiz questions to make them more difficult and added questions for the present study in 2017 that were intended to be more challenging. Based on student feedback, we also removed a few quiz questions that students interpreted differently than we intended. Final versions of quiz questions were reviewed by a member of the research team to check for clarity before deployment.

Survey creation

In the survey that we administered in week four and week eight, we included a Likert-scale question about the extent to which students perceived the videos influenced their learning (1 = *no influence* to 5 = *strongly influenced*), and a follow-up open-ended question asking the students to explain their reasoning for their response. Students were directed to answer only about the videos (monologue or dialogue) that they had watched most recently. The week eight survey also contained 5 Likert-scale questions measuring student-perceived effort while watching the videos in the past 4 weeks. The items were adapted from the Intrinsic Motivation Inventory (IMI) (Ryan, 1982). The students were asked to rate the items based on their experience for the past 4 weeks. Each question was rated from 1 (*not at all true*) to 7 (*very true*). The construct validity of the survey was provided by the developer, and the reliability of the survey used in the current study was at a good level (Cronbach's $\alpha = .89$).

Students' learning with the error episodes in the dialogue-videos

Identification of error episodes and the corresponding quiz questions

Students watched videos without segmenting, but for research analysis we segmented each video into two to six episodes based on the number of problems presented in the video (both monologue and dialogue). The error episodes were those that contain tutees suggesting an incorrect solution to a problem in the dialogue-videos. One coder with an expertise in physiology reviewed all of the dialogue-videos and out of the 36 total episodes, identified 13 error episodes. Then, two coders, both with expertise in physiology,

independently coded the quiz questions that measured the content featured in each error episodes and discussed any discrepancies until they reached an agreement on 37 quiz questions that measured the content featured in the 13 error episodes. The initial inter-rater reliability (Cohen’s κ) was .82, which indicated a good inter-rater reliability (Landis & Koch, 1977).

Data analysis

To compare student learning from dialogue- and monologue-videos, we first needed to reorganise students’ scores from the quizzes. That is, students’ learning from monologue-videos was measured by quizzes from when group A students watched the monologue-videos (during weeks one to four) and when group B students watched the monologue-videos (during weeks five to eight). Students’ learning from dialogue-videos was measured by quizzes completed by group B students from week one to week four and group A students from week five to week eight (Figure 2).

Analyses were performed in the IBM Statistical Package for the Social Sciences (SPSS) 25. To compare the observing student learning from error episodes in the dialogue-videos and the corresponding episodes in the monologue-videos, an independent sample *t*-test was carried out. Moreover, given the fact that prior knowledge can influence student learning from the error episodes (Arguel & Lane, 2015). To test out to what extent the impact of error episodes on student learning was affected by their GPA, a moderation analysis was carried out by using Process Macro 3.4 developed by Hayes (2013) and implemented for SPSS.

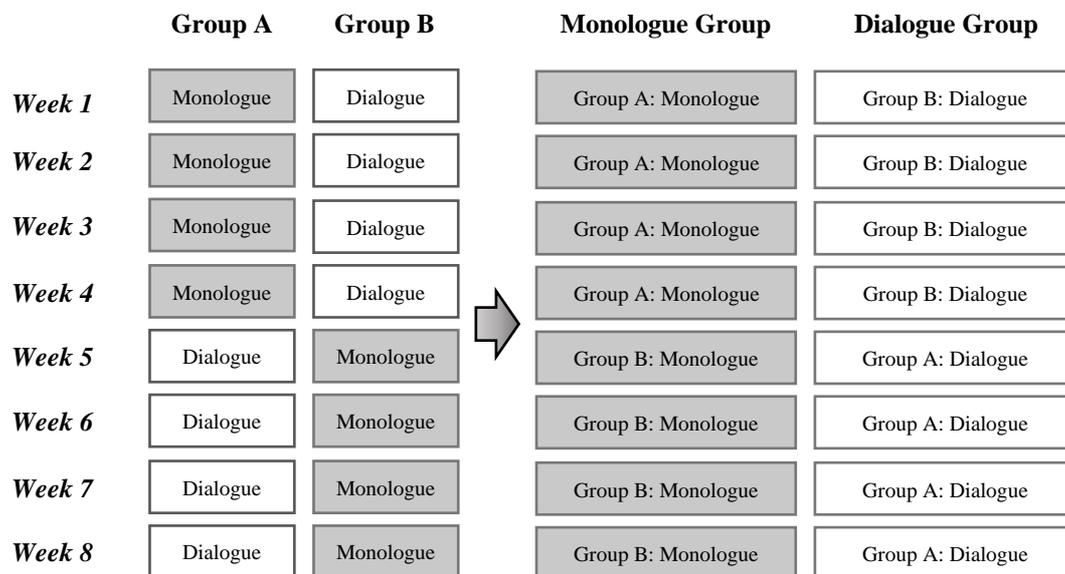


Figure 2. An illustration of counterbalancing design and regrouping for data analysis

Results

There was no significant difference in the scores on error episode-related quiz questions between students who watched the error episodes in dialogue-videos and the scores of students who watched the corresponding monologue episodes without including GPA in the model ($t[432] = -1.17, p = .244$). When we included GPA in the model, the moderation analysis again showed no significant main interaction in the overall model, where $b = .04, t[430] = 1.52$ with $p = .129 > .05$. However, a further probing analysis by using Johnson-Neyman technique revealed that student GPA began to affect their learning from the error episodes when student GPA is at least 3.42 ($t[430] = 1.97, p = .05, b = .025$). As GPA increases, the relationship between student learning and the conditions of watching error episodes from dialogue-videos or the corresponding episodes in monologue-videos became more positive with the highest GPA (4.23) ($b = .058, t[430] = 2.34, p = .02 < .05$). As shown in the Table 4, students whose GPA was equal to or higher than 3.42 learned better from the error episodes in dialogue-videos compared to the same episodes without student tutees’ errors in monologue-videos.

Table 4
Descriptive statistics for student performance on quiz questions related to error episodes

	N	Monologue		Dialogue	
		M	SD	M	SD
Overall comparison	217	.801	.146	.818	.153
GPA \geq 3.42	120	.842	.130	.881	.111
GPA $<$ 3.42	97	.760	.159	.755	.166

Note. Maximum score is 1

Factors affect learning from error episodes

Hereafter we use the term *higher-performers* to represent students whose GPA was equal to or higher than 3.42, and students whose GPA was lower than 3.42 are called *lower-performers*. To further investigate what factors may have contributed to the higher-performing students with relatively higher GPA performed better on the post-tests from the error episodes, whereas this pattern did not hold for relatively lower-performing students. We specifically analysed the relevant questions that related to the effort spent on watching dialogue-videos and the students' perceived influence of dialogue-videos asked in a survey. This survey contained several subscales and included open-ended questions asking their preference of dialogue- or monologue-videos and was administered at the end of week eight. In addition, we hypothesised that higher-performing students may have been more likely to engage in deep learning by completing the worksheets and consequently being more cognitively engaged when watching the dialogue-videos. We, therefore, also compared the worksheet completion rate of higher-performers with lower-performers. More specific information is provided below.

Data analysis

To find out whether effort was a factor that influenced students' learning from error-episodes, we pooled all students who watched dialogue-videos in the past 4 weeks (group A) from the week eight survey, and compared the higher-performers ($n = 68$) and lower-performers ($n = 42$) perceived effort. An independent sample *t*-test was carried out to compare to what extent there were differences in the effort that the lower-performing students and higher-performing students spent on watching the dialogue-videos. For the worksheet completion comparison between the higher-performing and lower-performing students, a binary score system was used to indicate the completion rate. Students who completed all the questions on the worksheet received one point for that week, and all incomplete and blank worksheets were scored as zero; each student had four binary scores for the four worksheets corresponding to the 4 weeks of dialogue-videos. We then calculated a percentage of worksheet completion rate for each student across the 4 weeks. An independent *t*-test was carried out to compare the higher-performing students' and lower-performing students' completion rate.

For perceived influence, we wanted to explore how higher-performing students and lower-performing students perceived dialogue-videos differently in assisting their learning. Therefore, we only pooled all student responses about the dialogue-videos ($n_{\text{higher}} = 94$; $n_{\text{lower}} = 73$), and only focused on the responses that perceived the dialogue-videos helped their learning (i.e., who responded to the Likert-scale item as 3 or higher). The analysis of the follow-up open-ended question consisted of three steps. First, one researcher reviewed all student responses to the open-ended survey question. An open-coding method was first used to identify common themes in student responses (Strauss & Corbin, 1998), and then the researcher constantly compared each theme to ensure that the description of the theme was inclusive of all responses (Glesne & Peshkin, 1992). Second, to test the reliability of the developed themes, the researcher and another researcher independently coded 20% of students responses by using the themes and their inter-rater score was at an acceptable level ($\kappa = .81$) (Landis & Koch, 1977). Third, Chi-square tests of independence were conducted to test whether there were differences in the percent of higher- and lower-performing students who reported a particular category. Some categories were reported by too few students to warrant meaningful interpretation of a Chi-square tests, and therefore tests were not performed.

Results

No significant difference was found on student perceived effort spent on watching dialogue-videos between the higher- and lower-performers ($t[108] = .40, p > .05$). On average higher-performers rated 4.79 ($SD = 1.46$), and lower-performers rated 4.90 ($SD = 1.28$) on the scales. Similarly, there was no significant difference of the worksheet attempt between the higher-performers and lower-performers ($t[215] = -1.57, p > .05$), although the higher-performing students had a slightly higher worksheet attempt ($M = .92, SD = .13$) compared to the lower-performing students ($M = .89, SD = .17$).

Students highlighted an array of ways that they perceived the dialogue-videos positively influenced their learning. Students reported that the dialogue-videos influenced their learning because they enhanced students' understanding of the content, focused students' attention on what content was important, provided an alternative way of learning, helped students engage in the content, and helped students prepare for class. Nevertheless, Chi-square tests revealed no statistically significant differences between the percent of higher- and lower-performing students who reported these themes. However, a significantly higher percent of higher-performing than lower-performing students reported that the dialogue-videos repeated the content that had already been taught in class, so that they watched the dialogue-videos for the purpose of reviewing the material ($\chi^2 = 7.10, p < 0.01$).

Discussion

Compared to the corresponding episodes in monologue-videos, we found that higher-performing students performed better on quizzes after watching episodes of the dialogue-videos where tutees made mistakes followed by the instructor's corrections and clarifications. However, no such pattern was found for lower-performing students. Moreover, significantly more higher-performing students treated the videos as reviewing materials than lower-performing students. This finding suggests that the higher-performing students may have established a stronger foundation of physiology knowledge than lower-performing students and were therefore better equipped to deal with error episodes in the dialogue-videos and the resulting confusion. Furthermore, this finding indicates that the cognitive disequilibrium that dialogue-videos may have caused, did not hinder higher-performing students' understanding of the content covered in the videos. The findings of our study have some direct implications for how to implement dialogue-videos in a real course to benefit student learning and the design of dialogue-videos, and the results of this study prompt ideas for future research.

Establishing necessary prior knowledge before watching dialogue-video

Taking the findings of this study altogether, it suggests that the errors made by tutees in dialogue-videos supported student understanding of physiology concepts, but only when students had the necessary knowledge base to resolve their confusion. When encountering confusion, higher-performing students are more likely to have the sufficient prior knowledge and/or well-developed learning strategies to conduct effective learning activities to timely solve the confusion, whereas lower-performing students may lack the ability to address the confusion by themselves if there is no scaffolding provided (Arguel et al., 2017). Therefore, to maximise the benefits of dialogue-video and to enable lower-performing students to also benefit from dialogue-videos, it is critical to prepare students with necessary prior knowledge specifically pertaining to the content covered in the videos. For example, this can be done by assigning readings, providing course notes for core concepts, or a short monologue-video.

Encouraging online collaborations when confusion cannot be resolved alone

Research has shown that confusion can actually inspire greater depth of cognitive processing, however, prolonged confusion will deter students from executing effort to regulate the cognitive disequilibrium (Craig et al., 2004; D'Mello et al., 2014). In a prior laboratory-based study demonstrating that students learned better when watching dialogue-videos compared to monologue-videos, students watched the videos in dyads and had the opportunity to discuss any error episodes, or points of confusion, with their partners (Muldner et al., 2014). However, students watched the videos individually in our study. When confusion occurred, there was no direct channel for the students to immediately resolve it. That could explain why higher-performing students benefited from error episodes in dialogue-videos, whereas no benefit was found

for lower-performing students. A possible solution to overcome this issue can be including an asynchronous online help-seeking forum (Ding & Er, 2017). Therefore, when confusion occurs, lower-performing students will have peers or course instructors to discuss and to resolve the confusion.

Embedding a confusion detection mechanism when watching dialogue-video

It could also be possible that lower-performing students simply give up trying to learn if they get confused (Arguel & Lane, 2015). If lower-performing students are more likely to avoid the learning opportunity from confusion than higher-performing students, then the higher-performing students may be the only ones benefitting from the confusion. Therefore, some confusion detection mechanism can be implemented to evaluate students' emotional statuses and to encourage students to continue to persist and take further actions to resolve any confusion. For example, Graesser et al. (2006) used three ways to detect student confusion: self-report, peer-report, and judge-report. It could be challenging for a peer-report or a judge-report to detect student confusion when watching dialogue-videos individually, but self-report can be easily implemented. For instance, embedding quiz questions in a video right after an error is made by a tutee to check whether students have fully understood the concept discussed between the instructor and the tutee featured in the video. If the answer is wrong, follow-up feedback can be provided by a system to encourage students to work out the confusion if there is any, and additional recourses can be provided to the student by the system to assist in resolving the confusion.

Aligning errors in dialogue-video with student understanding level

The results of this study show that the error episodes in dialogue-videos actually assisted learning for higher-performing students. However, the four student tutees recruited in our study who filmed the dialogue-videos completed the course in the previous year. Three of them received As and one received a B for their final grade in the course, so most would be considered higher-performers. Therefore, it could be possible that the tutees made fewer errors than the observing students would when asked the same question. In future research, to maximise the benefits of dialogue-videos, students whose academic abilities are more reflective of the average student in the course could be recruited for the videos. This way, more error episodes would be made, and observing students may identify themselves more with the tutees in the videos.

Further, it also could be possible that the tutees' errors were more sophisticated, and harder for a lower-performing student to comprehend. In future research, to understand how to maximise the advantages of dialogue-videos, researchers could investigate how the complexity of the type of error affects students. That is, are errors that the majority of students are likely to make more effective for all student learning compared to errors that only higher-performing students are likely to make? Students who have never taken the course and are not familiar with the topics in the course, or even current students who are struggling with the material, could be recruited as tutees for the videos to promote richer interaction that better matches the mastery levels of the observing students enrolled in the course.

Limitations

This study has several limitations. The primary concern of this study is that we do not have a way to confirm whether students actually watched the videos, and how long they spent on watching the videos. No instructional material can be beneficial if students do not utilise it. Due to the lack of this log data, the analyses for student learning outcomes were based on our assumption that all students watched the videos. In addition, given that the dialogue-videos were longer than the monologue-videos, especially the error-episodes that contain dialogues between the instructor and the student tutees, it could be possible that the additional time of error-episodes compared to the corresponding episodes in monologue would have differentially benefited higher-performing students. In contrast, the longer length of dialogue-videos may have had impaired student understanding of physiology concepts compared to monologue-videos given the fact that numerous studies have shown that shorter videos are more preferable (Scagnoli et al., 2015). If so, the error episodes in dialogue-videos could have even higher positive effect on student learning if the dialogue-videos were made equal length as the monologue-videos. A third concern is how we measured effort. We only included Likert-scaled items for measuring students' effort while watching dialogue-videos, which could only provide information for perceived effort. Moreover, the items were only included in the week eight survey. That is, the analysis was only based on half of the students' data.

Conclusion

The primary goal of this paper was to investigate if the error episodes in the dialogue-videos were beneficial or detrimental for observing student learning. We found that students with higher GPA performed better on the quizzes after watching error episodes in dialogue-videos compared to the corresponding episodes in monologue-videos; however, no such difference was found among lower performing students. The same pattern was found in student responses to the open-ended question focused on perceived benefits of dialogue-videos. We proposed that the error episodes in dialogue-videos can benefit student learning; however, only when students have already established a certain level of prior knowledge and sufficient self-regulated learning skills.

Acknowledgments

This study was funded by a National Science Foundation IUSE award #1504893 awarded to Michelene Chi and Sara Brownell. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We acknowledge the help of Amy Pate, Christiana Bruchok, David Yaghmourian, Joshua Adams, and Natalie Newton. We thank Logan Gin and Rachel Scott for their feedback on an earlier draft of this manuscript. Finally, we thank the students in the physiology course who took the time to provide feedback so that we can use evidence to make instructional decisions to enhance their experience and our anonymous reviewers for their careful and constructive reviews of prior drafts of this article.

References

- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101–128. <http://iospress.metapress.com/content/1qd3jqqtq69w9t1f>
- Arguel, A., & Lane, R. (2015). Fostering deep understanding in geography by inducing and managing confusion: An online learning approach. In T. Reiners, B. R. von Kinsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), *Globally connected, digitally enabled. Proceedings ascilite 2015* (pp. 22–26). Australasian Society for Computers in Learning in Tertiary Education.
- Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside out: Detecting learners' confusion to improve interactive digital learning environments. *Journal of Educational Computing Research*, 55(4), 526–551. <https://doi.org/10.1177/0735633116674732>
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students “Game the system.” *Proceedings of the SIGCHI conference on Human factors in computing systems*, 383–390. <https://doi.org/10.1145/985692.985741>
- Baker, R. S., D’Mello, S., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies*, 68(4), 223–241. <https://doi.org/10.1016/j.ijhcs.2009.12.003>
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301–341. <https://doi.org/10.1080/03640210701863396>
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533. [https://doi.org/10.1016/S0364-0213\(01\)00044-1](https://doi.org/10.1016/S0364-0213(01)00044-1)
- 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>
- Cooper, K. M., Ding, L., Stephens, M. D., Chi, M. T. H., & Brownell, S. E. (2018). A course-embedded comparison of instructor-generated videos of either an instructor alone or an instructor and a student. *CBE—Life Sciences Education*, 17(2), 1–15. <https://doi.org/10.1187/cbe.17-12-0288>

- Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241–250. <https://doi.org/10.1080/1358165042000283101>
- D’Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*, 25(7), 1299–1308. <https://doi.org/10.1080/02699931.2011.613668>
- D’Mello, S., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. <https://doi.org/10.1016/j.learninstruc.2011.10.001>
- D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. C. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170. <https://doi.org/10.1016/j.learninstruc.2012.05.003>
- Ding, L., Adams, J., Stephens, M., Brownell, S. E., & Chi, M. T. H. (2018). Failure to replicate using dialogue videos in learning: Lessons learned from an authentic course. *Proceedings of the 13th International Conference of the Learning Sciences: Rethinking Learning in the Digital Age: Making the Learning Sciences Count*, 953–956. <https://doi.org/10.22318/csc2018.953>
- Ding, L., & Er, E. (2017). Determinants of college students’ use of online collaborative help-seeking tools. *Journal of Computer Assisted Learning*, 34(2), 1–11. <https://doi.org/10.1111/jcal.12221>
- Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2004). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, 29(4), 431–450. <https://doi.org/10.2190/Q8CM-FH7L-6HJU-DT9W>
- Evans, H. K., & Cordova, V. (2015). Lecture videos in online courses: A follow-up. *Journal of Political Science Education*, 11(4), 472–482. <https://doi.org/10.1080/15512169.2015.1069198>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Fyfield, M., Henderson, M., Heinrich, E., & Redmond, P. (2019). Videos in higher education: Making the most of a good thing. *Australasian Journal of Educational Technology*, 35(5), 1–7. <https://doi.org/10.14742/ajet.5930>
- Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. Longman.
- Graesser, A. C., Chipman, P., King, B., McDaniel, B., & D’Mello, S. (2007). Emotions and learning with AutoTutor. *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 569–571.
- Graesser, A. C., Witherspoon, A., McDaniel, B., D’Mello, S. K., Chipman, P., & Gholson, B. (2006). Detection of emotions during learning with AutoTutor. In R. Son (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 285–290). Erlbaum.
- Hayes, A. F. (2013). *Mediation, moderation, and conditional process analysis*. Guilford Press.
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363–374. <https://doi.org/http://dx.doi.org/10.2307/2529786>
- Lehman, B., D’Mello, S., & Graesser, A. C. (2012). Confusion and complex learning during interactions with computer learning environments. *Internet and Higher Education*, 15(3), 184–194. <https://doi.org/10.1016/j.iheduc.2012.01.002>
- Lehman, B., D’Mello, S., & Person, N. (2010). The intricate dance between cognition and emotion during expert tutoring. In J. Kay, & V. Alevan (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 433–442). Springer. https://doi.org/10.1007/978-3-642-13437-1_1
- McDaniel, M. A., Mestre, J. P., Frey, R. F., Gouravajhala, R., Hilborn, R. C., Miyatsu, T., & Yuan, H. (2017). *Maximizing undergraduate STEM learning: Promoting research at the intersection of cognitive psychology and discipline-based education research*. <https://circle.wustl.edu/white-paper-maximizing-undergraduate-stem-learning/>
- Mestre, J. P., Chevillat, A., & Herman, G. L. (2018). Promoting DBER-cognitive psychology collaborations in STEM education. *Journal of Engineering Education*, 107(1), 5–10. <https://doi.org/doi:10.1002/jee.20188>
- Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, 106(1), 69–85. <https://doi.org/10.1037/a0034448>
- 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. <https://doi.org/10.1111/j.1365-2729.2007.00248.x>

- Muller, D. A., & Sharma, M. D. (2007). Tackling misconceptions in introductory physics using multimedia presentations. *Proceedings of UniServe Science Teaching and Learning Research*, 58–63. http://sydney.edu.au/science/uniserve_science/pubs/procs/2007/14.pdf%5Cnpapers3://publication/uuid/F7AF2B22-1B69-47AE-B79C-5BC988750801
- Muller, D. A., Sharma, M. D., Eklund, J., & Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science*, 35(6), 519–533. <https://doi.org/10.1007/s11251-007-9017-6>
- Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3), 531–549. <https://doi.org/10.1037/a0019243>
- Ryan, A. M., Patrick, H., & Shim, S. O. (2005). Differential profiles of students identified by their teacher as having avoidant, appropriate, or dependent help-seeking tendencies in the classroom. *Journal of Educational Psychology*, 97(2), 275–285. <https://doi.org/10.1037/0022-0663.97.2.275>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461. <https://doi.org/10.1037/0022-3514.43.3.450>
- Scagnoli, N. I., Choo, J., & Tian, J. (2017). Students' insights on the use of video lectures in online classes. *British Journal of Educational Technology*, 50(1), 399–414. <https://doi.org/10.1111/bjet.12572>
- Scagnoli, N. I., McKinney, A., & Moore-Reynen, J. (2015). Video lectures in e-learning. In F. Nafukho, & B. Irby (Eds.), *Handbook of research on innovative technology integration in higher education* (pp. 115–134). Information Science Reference. <https://doi.org/10.4018/978-1-4666-8170-5.ch006>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin. <http://impact.cgiar.org/pdf/147.pdf>
- Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Sage Publications.
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- 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>
- Wood, W. B., & Tanner, K. D. (2012). The role of the lecturer as tutor: Doing what effective tutors do in a large lecture class. *CBE—Life Sciences Education*, 11(1), 3–9. <https://doi.org/10.1187/cbe.11-12-0110>
- Zeidner, M. (2007). Test anxiety: conceptions, findings, conclusions. In P. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). Academic Press.
- Zohar, A., & Aharon-Kravetsky, S. (2005). Exploring the effects of cognitive conflict and direct teaching for students of different academic levels. *Journal of Research in Science Teaching*, 42(7), 829–855. <https://doi.org/10.1002/tea.20075>

Corresponding author: Lu Ding, lding@eiu.edu/lu.ding@msn.com

Copyright: Articles published in the Australasian Journal of Educational Technology (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: Ding, L., Cooper, K. M., Stephens, M. D., Chi, M. T. H., & Brownell, S. E. (2021). Learning from error episodes in dialogue-videos: The influence of prior knowledge. *Australasian Journal of Educational Technology*, 37(4), 20–32. <https://doi.org/10.14742/ajet.6239>