

Exploring the Impact of Different Types of Instructor Generated Videos on Student Learning in a University Physiology Course

Katelyn M. Cooper, Lu Ding, Michelle D. Stephens, Michelene T.H. Chi, Sara E. Brownell
Arizona State University

kmcoope1@asu.edu, luding@asu.edu, mstephens@asu.edu, michelene.chi@asu.edu, sara.brownell@asu.edu

Abstract: Videos have become a popular way to engage students with material prior to a class, yet this is an unexplored area of research. There is support for the use of videos where instructors tutor students, but few studies have been conducted in a classroom. In this study, we used a randomized crossover design to compare the impact of two types of instructor-generated videos on university students in a large-enrollment physiology course. We compared videos featuring only an instructor with videos featuring an instructor tutoring a student. We found that students preferred, enjoyed, and valued the Instructor Only videos significantly more than the Instructor-Tutee videos. Additionally, below-average students performed significantly better on physiology quizzes after watching Instructor Only videos compared with Instructor-Tutee videos. Above average students performed equivalently on physiology quizzes after watching Instructor Only or Instructor-Tutee videos. This study applies cognitive science to a classroom practice.

Introduction

College classes are increasingly transitioning from traditional lecture to active learning, where students engage in constructing their own knowledge during class (AAAS, 2011). One way of creating active learning classrooms is to “flip” the class so that students are expected to learn content on their own before class and then apply the concepts during in-class active-learning activities where they solve complex problems and work with peers (Brame, 2013).

In flipped classrooms, students learn content on their own before class by reading the textbook, reading articles, watching animations, or viewing recorded lectures. Few studies have compared how different methods of content delivery may affect students’ attitudes and performance. However, in one study of students enrolled in an introductory college biology class, researchers used a two-by-two study design and compared the effects of video pre-class assignments compared to textbook pre-class assignments. Students who were assigned to watch the videos were significantly more satisfied with their pre-class assignment than students who were assigned to read the textbook. This study and others have led to recommendations for using videos as an educational tool to help prepare students for class (Rackaway, 2012; Stockwell *et al.*, 2015). However, if an instructor is interested in using videos as part of a pre-class assignment, how should these videos be designed and implemented?

In the literature, tutees learning from tutors is considered the gold standard, in that it exceeds all other forms of instruction in helping students achieve learning gains (Bloom, 1984; VanLehn, 2011). Tutoring has been shown to have effect sizes ranging from 0.79 to 2.0 S.D (Bloom, 1984; VanLehn, 2011). Based on this literature, Chi and colleagues (2008) have argued that capturing tutor-tutee dialog in videos (Instructor-Tutee videos) and re-using it for students to watch would result in greater learning for the student watching the videos (observing student) than only watching a video of an instructor alone (Instructor Only videos).

Cognitive science studies have supported the hypothesis that students learn from watching other students being tutored. For example, in two separate studies, observing students who were watching a video to learn about a particular topic learned as well as the individual tutees in the videos; this result was demonstrated in two difficult STEM domains: learning to solve physics problems and learning to explain diffusion in chemistry (Chi *et al.*, 2008; Muldner *et al.*, 2014). Additionally, several other laboratory studies have found that, for observing students, videos where a student is being tutored can be more beneficial for learning than videos featuring only an instructor (Driscoll *et al.*, 2003; Fox Tree, 1999). While these findings suggest that instructors should create Instructor-Tutee videos to enhance student learning, these prior studies were situated in a controlled laboratory setting that was removed from the context of a real class. To our knowledge, no studies have been conducted over consecutive weeks in a college classroom. However, there have been recent calls for more collaborations between cognitive scientists and discipline-based education researchers to see if lab-based cognitive science findings can be replicated in a formal class environment and to see what additional information we can glean from applying these theories to real classrooms (McDaniel *et al.*, 2017; Mestre *et al.*, in press).

The primary goal of this study was to compare the impact of Instructor-Tutee videos and Instructor Only videos on student performance in the context of a large-enrollment upper-division, active-learning, college physiology

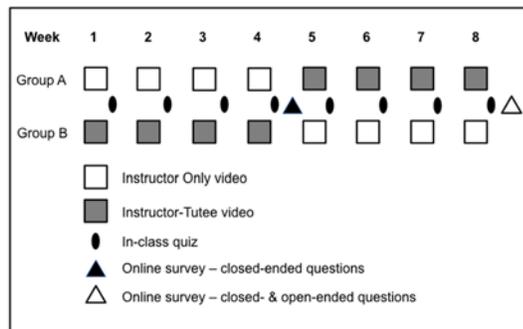


Figure 2. Depiction of experimental design

Instructor-Tutee videos

In the Instructor-Tutee videos, the instructor tutored a student during the video. All four students who appeared in the videos (tutees from here forward) had completed the same physiology course in fall 2015, and thus, were familiar with the content of the course. Only one tutee interacted with the instructor in each video. There were eight Instructor-Tutee videos, one for each week of the experiment. Each of the four tutees appeared in two of the eight videos, so that the observing students in both Group A and Group B watched a video with each of the four tutees.

In the Instructor-Tutee videos, the instructor would pose a physiology question and allow a few minutes for the tutee and the physiology students watching the videos (hereafter observing students) to think about how to answer the question. The tutee would then attempt to solve the problem and the instructor would ask guiding questions so that the student fully elaborated on his or her thoughts. The tutees attempted five to seven physiology problems in each video. After the tutee attempted a problem, the instructor corrected any misconceptions brought up by the tutee and articulated or elaborated on the correct solution to the physiology problem. The Instructor-Tutee videos ranged from 16 minutes – 27 minutes and averaged 21 minutes.

Instructor Only videos

Instructor Only videos featured only the instructor of the course. Similar to the Instructor-Tutee videos, the instructor would guide the observing students through the same problems that were presented in that week's Instructor-Tutee videos and give the observing students a few minutes of think-time to consider how to answer a problem before working through the problem by writing and talking out the answer. The Instructor Only videos ranged from 12 minutes – 21 minutes, and averaged 17 minutes. Independent t-test revealed that the Instructor Only videos were significantly shorter than the Instructor-Tutee videos ($t=2.3393$, $p=0.0347$).

Survey of opinions about the videos

A survey consisting of Likert-scale questions was administered to students after week four and a survey consisting of Likert-scale and open-ended questions was administered to students after week eight of the course. Both surveys asked students about their perceived usefulness of the videos and the extent to which they enjoyed watching the videos. Five Likert-scale items measuring student perceived usefulness and five Likert-scale items measuring perceived enjoyment of watching the videos were adapted from the Intrinsic Motivation Inventory (IMI; Ryan, 1982); each of which was rated from 1 (not at all true) to 7 (very true). Items were slightly reworded to reflect the video context of this study. Reliabilities (Cronbach's α) of the two constructs in the current study were at an acceptable level. The week eight survey asked students which video they preferred and asked them to explain why.

Analyses

Three authors reviewed all students' reasoning why they preferred either Instructor Only or Instructor-Tutee videos. Using open coding methods, the authors identified common themes in student responses and created a rubric to code each question (Strauss and Corbin, 1990). Then two authors independently reviewed 25% of student responses. Their inter-rater reliability was 94% for the question about why students prefer Instructor Only videos and 98% for why students prefer Instructor-Tutee videos. One author coded the remaining responses for each question.

To compare the effect of Instructor Only videos and Instructor-Tutee videos on the students' reported usefulness, enjoyment, and their performance, the data from student surveys about their perceived usefulness and enjoyment of each type of video, as well as their performance on all eight video quizzes, were reorganized into two datasets: Instructor Only and Instructor-Tutee. That is, Group A survey data from weeks 1-4 and Group B survey data

from weeks 5-8 were combined, and renamed as Instructor Only video survey data. Group A survey data from weeks 5-8 and Group B survey data from weeks 1-4 were combined, and renamed as Instructor-Tutee video survey data. The same reorganization was applied to the students' performance data on the physiology quizzes.

Before conducting formal analyses, datasets were screened and modified for missing values. Little's MCAR test (1988) was performed, and the results indicated that the missing values in the datasets were missing completely at random. Multiple imputation (MI) was used to impute missing values because it provides robust estimates when the data is missing completely at random (Schafer and Graham, 2002).

The assumption of normality was violated for paired sample t-tests and thus Wilcoxon signed-rank test was conducted to compare student perceived usefulness and enjoyment of the videos, as well as to compare student performance on the quizzes. Furthermore, research has shown that compared to high performing students, low performing students tended to benefit more from engaged learning activities (Carini *et al.*, 2006; Freeman *et al.*, 2011). Therefore, we divided students into two groups- lower achieving students with a prior grade point average (GPA) below the median course prior GPA (GPA < 3.49) and higher achieving students with a GPA at or above the median (GPA ≥ 3.49). Two Wilcoxon signed-rank tests were carried out to examine student learning gains.

Results

Finding 1: The majority of students prefer Instructor Only videos

In response to a Likert-scale survey question, we found that the majority of students (59.9%) preferred the Instructor Only videos, while 20.3% of students preferred the Instructor-Tutee videos and 19.8% of students reported that they did not prefer one type of video over the other. These frequencies were significantly different ($\chi^2(2, 207) = 65.77, p = 0.00 < 0.05$). We found that students preferred Instructor Only videos because they perceived that they understood the content better when watching the Instructor Only videos and they perceived the Instructor-Tutee videos to be confusing. Specifically, they found the Instructor Only videos to be straightforward and the Instructor-Tutee videos to be indirect. They also highlighted that Instructor only videos presented only correct information, while the tutees in the Instructor-Tutee Videos sometimes provided incorrect information, which they perceived to be confusing. Lastly, students highlighted that Instructor Only videos were shorter than Instructor-Tutee videos.

Finding 2: Students perceived higher usefulness and enjoyment with regard to watching Instructor Only videos compared with Instructor-Tutee videos

Students perceived higher usefulness ($Z = -2.61, p = 0.013, \text{Cohen's } d = 0.18$) and enjoyment level ($Z = -3.69, p = 0.000, \text{Cohen's } d = 0.25$), with regard to watching the Instructor Only videos compared with the Instructor-Tutee videos. However, effect sizes are small. **Table 1** presents descriptive statistics and Wilcoxon signed-rank test results for the analyses of student perceived usefulness and enjoyment of the different types of videos.

Table 1: Descriptive Statistics and Wilcoxon Signed-Rank Test Results for Survey Data

* Holm-Bonferroni adjustment applied on the α level.

	N	Instructor Only videos		Instructor-Tutee videos		Z	p	df
		M	SD	M	SD			
Perceived usefulness	217	5.48	1.26	5.21	1.26	-2.61	.013*	216
Perceived enjoyment	217	4.62	1.31	4.27	1.36	-3.69	.000*	216

Finding 3: Students with a GPA below the median perform better after watching Instructor Only videos compared with Instructor-Tutee videos

The results of Wilcoxon signed-rank test revealed no significant difference in students' quiz scores after they watched Instructor Only videos (M = 7.95, SD = 1.04, N = 217) and Instructor-Tutee videos (M = 7.89, SD = 1.00, N = 217). However, results of further analyses after disaggregating students into a lower achieving group (GPA < 3.49, N = 107) and a higher achieving group (GPA ≥ 3.49, N = 110) revealed that students in the lower achieving group performed significantly better after watching Instructor Only videos (M = 7.58, SD = 1.03) compared to Instructor-Tutee videos (M = 7.36, SD = 0.99) with a small effect size (Cohen's $d = 0.25$). No significant difference in student performance after watching the Instructor Only or Instructor-Tutee videos was found within the higher achieving group of students. **Table 2** shows descriptive statistics and Wilcoxon signed-rank test results for the analyses conducted on student performance data.

Table 2. Descriptive Statistics and Wilcoxon Signed-Rank Test Results for Student Performance.

* Holm-Bonferroni adjustment applied on the α level.

	N	Instructor Only videos		Instructor-Tutee videos		Z	p	df
		M	SD	M	SD			
Overall comparison	217	7.95	1.04	7.89	1.00	-1.26	.214	216
Higher achieving group	110	8.32	.71	8.41	.65	-.97	.348	109
Lower achieving group	107	7.58	1.03	7.34	.99	-2.75	.008*	106

Discussion

While the previous lab-based studies showed that students performed better after watching a video of an instructor tutoring another student compared to watching a video of only an instructor, we did not find that pattern in this study. One interpretation for the conflicting results is that the cognitive science lab-based studies did not account for some of the factors that influence students in a real college course. While lab-based studies have the advantage of being reductionist and often being able to control for many factors, they often lack the complexities of a real classroom setting. One assumption of this study is that students watched and fully engaged with the videos. In the context of a lab-based study, students have no other distractions that may prevent them from fully engaging with the video. However, in the context of a real course, students have competing demands on their time, which may influence their engagement with the videos. Students highlighted they preferred the Instructor Only videos because of the short length. It is possible that something as simple as the difference in video length – as opposed to whether the instructor was tutoring a student – could result in a difference in engagement and may partially explain why lower performing students performed better after watching the shorter Instructor Only videos (Guo *et al.*, 2014).

We chose to include the same amount of content in both types of videos, which meant that the Instructor-Tutee videos had to be longer because of the time needed to interact with students. The decision to cover the same amount of content in both videos, and thus make the videos different lengths, was made because these videos covered course content that students would be tested on and we did not think that it was fair for some students to be exposed to content that other students would not be exposed to. It would be interesting to compare similar length videos in the future to try to establish whether the length of the video was an important factor.

Lower achieving students scored higher on quizzes after watching Instructor Only videos

We found that lower achieving students scored higher on quizzes after watching the Instructor Only videos compared to the Instructor-Tutee videos. This may be because the Instructor-Tutee videos exceeded the capacity of students' cognitive load. Working memories have a very limited capacity and thus, students have a limited capacity to process information (Mayer, 2008). Students in this study reported that when both an incorrect and a correct idea were provided during the video, they felt confused because they struggled to remember which idea was correct. This may indicate that the presentation of incorrect information, in addition to correct information about physiology, required extraneous processing that did not directly support students' learning and could be the underlying reason for why the lower achieving students performed worse after watching the Instructor-Tutee videos. We hypothesize that either lower achieving students have less working memory capacity, or perhaps higher achieving students spent extra time reviewing the videos in order to clarify what information was correct even if their working memory capacity was exceeded when watching the video for the first time.

Recommendations and limitations

This study was conducted in the context of one physiology course at one institution and a single instructor was present in every video. As we highlight in this manuscript, context matters and this limits the generalizations of our findings beyond this particular context. Further, our finding that lower achieving students perform better after watching Instructor Only videos compared with Instructor-Tutee videos contradicts previous lab studies and the significant difference in lower achieving students' performance on the physiology quiz is small. Based on these contradictory results, we cannot recommend either Instructor Only or Instructor-Tutee videos for the purpose of enhancing student performance at this time.

Conclusion

In this study, we explore student perceptions of and performance on physiology quizzes after watching Instructor Only videos and Instructor-Tutee videos. We found that students are more likely to prefer, value, and enjoy Instructor Only

videos. In contrast with previous lab studies, we also found that lower performing students performed better on physiology quizzes after watching Instructor Only videos compared with Instructor-Tutee videos.

References

- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC. Retrieved November 28, 2017, from <http://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf>.
- 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.
- Brame, C. J. (2013). Flipping the classroom. Retrieved February 8, 2018, from <https://cft.vanderbilt.edu/wp-content/uploads/sites/59/Flipping-the-classroom.pdf>.
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in higher education*, 47(1), 1-32.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, 32(2), 301-341.
- Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, 29(4), 431-450.
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse processes*, 27(1), 35-53.
- Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE-Life Sciences Education*, 10(2), 175-186.
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198-1202.
- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American psychologist*, 63(8), 760.
- 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. Retrieved from <https://circle.wustl.edu/wp-content/uploads/2017/06/McDaniel-et-al.-2017-cognitive-psychology-and-discipline-based-education-research.pdf>.
- Mestre, J.P., Cheville, A., & Herman, G. L. (in press). Promoting DBER-cognitive psychology collaborations in STEM education. *Journal of Engineering Education*.
- Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, 106(1), 69.
- Rackaway, C. (2012). Video killed the textbook star?: Use of multimedia supplements to enhance student learning. *Journal of Political Science Education*, 8(2), 189-200.
- 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.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147.
- Stockwell, B. R., Stockwell, M. S., Cennamo, M., & Jiang, E. (2015). Blended learning improves science education. *Cell*, 162(5), 933-936.
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.

Acknowledgements

This study was funded by NSF IUSE award #1504893.