How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos

Philip J. Guo
MIT CSAIL / University of Rochester
pg@cs.rochester.edu

Juho Kim
MIT CSAIL
juhokim@mit.edu

Rob Rubin
edX
rrubin@edx.org

ABSTRACT
Videos are a widely-used kind of resource for online learning. This paper presents an empirical study of how video production decisions affect student engagement in online educational videos. To our knowledge, ours is the largest-scale study of video engagement to date, using data from 6.9 million video watching sessions across four courses on the edX MOOC platform. We measure engagement by how long students are watching each video, and whether they attempt to answer post-video assessment problems.

Our main findings are that shorter videos are much more engaging, that informal talking-head videos are more engaging, that Khan-style tablet drawings are more engaging, that even high-quality pre-recorded classroom lectures might not make for engaging online videos, and that students engage differently with lecture and tutorial videos.

Based upon these quantitative findings and qualitative insights from interviews with edX staff, we developed a set of recommendations to help instructors and video producers take better advantage of the online video format. Finally, to enable researchers to reproduce and build upon our findings, we have made our anonymized video watching data set and analysis scripts public. To our knowledge, ours is one of the first public data sets on MOOC resource usage.

Author Keywords
Video engagement; online education; MOOC

ACM Classification Keywords
H.5.1. Information Interfaces and Presentation (e.g. HCI): Multimedia Information Systems

INTRODUCTION
Educators have been recording instructional videos for nearly as long as the format has existed. In the past decade, though, free online video hosting services such as YouTube have enabled people to disseminate instructional videos at scale. For example, Khan Academy videos have been viewed over 300 million times on YouTube [1].

L@S 2014, March 4–5, 2014, Atlanta, Georgia, USA.

ACM 978-1-4503-2669-8/14/03..$15.00.
http://dx.doi.org/10.1145/2556325.2566239

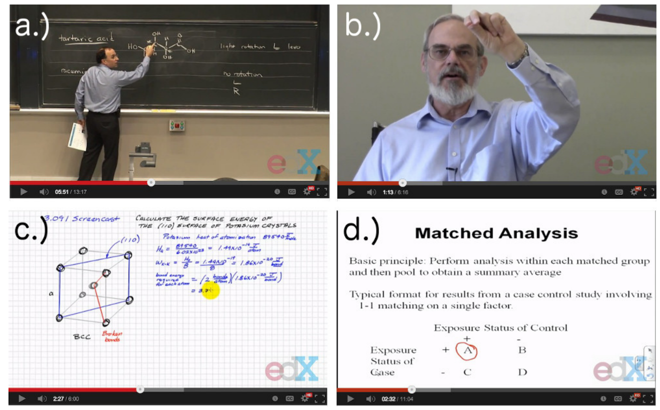

Figure 1. Video production style often affects student engagement in MOOCs. Typical styles include: a.) classroom lecture, b.) "talking head" shot of an instructor at a desk, c.) digital tablet drawing format popularized by Khan Academy, and d.) PowerPoint slide presentations.

Videos are central to the student learning experience in the current generation of MOOCs from providers such as Coursera, edX, and Udacity (sometimes called xMOOCs [7]). These online courses are mostly organized as sequences of instructor-produced videos interspersed with other resources such as assessment problems and interactive demos. A study of the first edX course (6.002x, Circuits and Electronics) found that students spent the majority of their time watching videos [2, 13]. Also, a study of three Coursera courses found that many students are auditors who engage primarily with videos while skipping over assessment problems, online discussions, and other interactive course components [9].

Due to the importance of video content in MOOCs, video production staff and instructional designers spend considerable time and money producing these videos, which are often filmed in diverse styles (see Figure 1). From our discussions with staff at edX, we learned that one of their most pressing questions was: *Which kinds of videos lead to the best student learning outcomes in a MOOC?* A related question that affects the rate at which new courses can be added is how to maximize student learning while keeping video production time and financial costs at reasonable levels.

As a step toward this goal, this paper presents an empirical study of students' *engagement* with MOOC videos, as measured by how long students are watching each video, and whether they attempt to answer post-video assessment problems. We choose to study engagement because it is a necessary (but not sufficient) prerequisite for learning, and because it can be quantified by retrospectively mining user interaction logs from past MOOC offerings. Also, video engagement

Finding	Recommendation
Shorter videos are much more engaging.	Invest heavily in pre-production lesson planning to segment videos into chunks shorter than 6 minutes.
Videos that intersperse an instructor's talking head with slides are more engaging than slides alone.	Invest in post-production editing to display the instructor's head at opportune times in the video.
Videos produced with a more personal feel could be more engaging than high-fidelity studio recordings.	Try filming in an informal setting; it might not be necessary to invest in big-budget studio productions.
Khan-style tablet drawing tutorials are more engaging than PowerPoint slides or code screencasts.	Introduce motion and continuous visual flow into tutorials, along with extemporaneous speaking.
Even high quality pre-recorded classroom lectures are not as engaging when chopped up for a MOOC.	If instructors insist on recording classroom lectures, they should still plan with the MOOC format in mind.
Videos where instructors speak fairly fast and with high enthusiasm are more engaging.	Coach instructors to bring out their enthusiasm and reassure that they do not need to purposely slow down.
Students engage differently with lecture and tutorial videos	For lectures, focus more on the first-watch experience; for tutorials, add support for rewatching and skimming.

Table 1. Summary of the main findings and video production recommendations that we present in this paper.

is important even beyond education. For instance, commercial video hosting providers such as YouTube and Wistia use engagement as a key metric for viewer satisfaction [6, 16], which directly drives revenues.

The importance of scale: MOOC video producers currently base their production decisions on anecdotes, folk wisdom, and best practices distilled from studies with at most dozens of subjects and hundreds of video watching sessions. The scale of data from MOOC interaction logs—hundreds of thousands of students from around the world and millions of video watching sessions—is four orders of magnitude larger than those available in prior studies [11, 15].

Such scale enables us to corroborate traditional video engagement research and extend their relevance to a modern online context. It also allows MOOC video producers to make more rigorous decisions based on data rather than just intuitions. Finally, it could enable our findings and recommendations to generalize beyond MOOCs to other sorts of informal online learning that occurs when, say, hundreds of millions of people watch YouTube how-to videos on topics ranging from cooking to knitting.

This paper makes three main contributions:

- **Findings** from an empirical study of MOOC video engagement, combining data analysis of 6.9 million video watching sessions in four edX courses with interviews with six edX production staff. The left column of Table 1 summarizes our seven main findings. To our knowledge, ours is the largest-scale study of video engagement to date.

- **Recommendations** for instructional designers and video producers, based on our study's findings (see the right column of Table 1). Staff at edX are already starting to use some of these recommendations to nudge professors toward cost-effective video production techniques that lead to greater student engagement.

- **An anonymized public data set** of 6.9 million video watching sessions, along with analysis scripts and installation instructions to enable full reproducibility of our results. Located at `http://www.pgbovine.net/edX/`, ours is one of the first public data sets on MOOC resource usage.

RELATED WORK

To our knowledge, our study is the first to correlate video production style with engagement at scale using millions of viewing sessions.

The closest related work is by Cross et al., who studied some of these effects in a controlled experiment [4]. They created Khan-style (tablet drawing) and PowerPoint slide versions of three video lectures and surveyed 150 people online about their preferences. They found that the two formats had complementary strengths and weaknesses, and developed a hybrid style called TypeRighting that tries to combine the benefits of both. Ilioudi et al. performed a similar study using three pairs of videos recorded in both live classroom lecture and Khan-style formats, like those shown in Figure 1a. and c., respectively. They presented those videos to 36 high school students, who showed a slight preference for classroom lecture videos over Khan-style videos [8]. Although these studies lack the scale of ours, they collected direct feedback from video watchers, which we have not yet done.

Prior large-scale analyses of MOOC interaction data (e.g., [2, 3, 9, 13]) have not focused on videos in particular. Some of this work provides the motivation for our study. For instance, a study of the first edX course (6.002x, Circuits and Electronics) found that students spent the majority of their time watching videos [2, 13]. And a study of three Coursera courses

Course	Subject	University	Lecture Setting	Videos	Students	Watching sessions
6.00x	Intro. CS & Programming	MIT	Office Desk	141	59,126	2,218,821
PH207x	Statistics for Public Health	Harvard	TV Studio	301	30,742	2,846,960
CS188.1x	Artificial Intelligence	Berkeley	Classroom	149	22,690	1,030,215
3.091x	Solid State Chemistry	MIT	Classroom	271	15,281	806,362
Total				862	127,839	6,902,358

Table 2. Overview of the Fall 2012 edX courses in our data set. "Lecture Setting" is the location where lecture videos were filmed. "Students" is the number of students who watched at least one video.

found that many students are auditors who engage primarily with videos while skipping over assessment problems, online discussions, and other interactive course components [9].

Finally, educators have been using videos and electronic media for decades before MOOCs launched. Mayer surveys cognitive science research on the impacts of multimedia on student learning [11]. Williams surveys general instructional media best practices from the 1950s to 1990s [15]. And Levasseur surveys best practices for using PowerPoint lectures in classrooms [10]. These studies have at most dozens of subjects and hundreds of video watching sessions. Our study extends these lines of work to a large-scale online setting.

METHODOLOGY

We took a mixed methods approach: We analyzed data from four edX courses and supplemented our quantitative findings with qualitative insights from interviews with six edX staff who were involved in producing those courses.

Course Selection

We analyzed data from four courses in the first edX batch offered in Fall 2012 (see Table 2). We selected courses from all three edX affiliates at the time (MIT, Harvard, and UC Berkeley) and strived to maximize diversity in subject matter and video production styles (see Figure 1).

However, since all Fall 2012 courses were math/science-focused, our corpus does not include any humanities or social science courses. EdX launched additional courses in Spring 2013, but that data was incomplete when we began this study. To improve external validity, we plan to replicate our experiments on more courses once we obtain their data.

Video Watching Sessions

The main data we analyze is a *video watching session*, which represents a single instance of a student watching a particular edX video. Each session contains a username, video ID, start and end times, video play speed (1x, 1.25x, 1.5x, 0.75x, or multiple speeds), numbers of times the student pressed the play and pause buttons, and whether the student attempted an assessment problem shortly after watching the given video.

To extract video watching sessions, we mined the edX server logs for our four target courses. The edX website logs user interaction events such as navigating to a page, playing a video, pausing a video, and submitting a problem for grading. We segmented the raw logs into video watching sessions based on these heuristics: Each session starts with a "play video" event for a particular student and video, and it ends when:

- that student triggers any event not related to the current video (e.g., navigating to another page),
- that student ends the current login session,
- there is at least a 30-minute gap before that student's next event (Google Analytics [5] uses this heuristic for segmenting website visits),
- the video finishes playing. The edX video player issues a "pause video" event when a video ends, so if a student plays, say, a five-minute video and then walks away from the computer, that watching session will conclude when the video ends after five minutes.

In Fall 2012, the edX video player automatically started playing each video (and issues a "play video" event) as soon as a student loads the enclosing page. Many students paused the video almost immediately or navigated to another page. Thus, we filtered out all sessions lasting shorter than five seconds, because those were likely due to auto-play.

Our script extracted 6.9 million total video watching sessions across four courses during the time period when they were initially offered in Fall 2012 (see Table 2).

Measuring Engagement

We aim to measure student engagement with instructional videos. However, true engagement is impossible to measure without direct observation and questioning, which is infeasible at scale. Thus, we use two proxies for engagement:

Engagement time: We use the length of time that a student spends on a video (i.e., video watching session length) as the main proxy for engagement. *Engagement time* is a standard metric used by both free video providers such as YouTube [6] and enterprise providers such as Wistia [16]. However, its inherent limitation is that it cannot capture whether a watcher is actively paying attention to the video or just playing it in the background while multitasking.

Problem attempt: 32% of the videos across our four courses are immediately followed by an assessment problem, which is usually a multiple-choice question designed to check a student's understanding of the video's contents. We record whether a student attempted the follow-up problem within 30 minutes after watching a video. A problem attempt indicates more engagement than moving on without attempting.

When we refer to *engagement* throughout this paper, we mean engagement as measured through these two proxies, not the difficult-to-measure ideal of true engagement.

Video Properties

To determine how video production correlates with engagement, we extracted four main properties from each video.

Length: Since all edX videos are hosted on YouTube, we wrote a script to get each video's length from YouTube.

Speaking rate: All edX videos come with time-coded subtitles, so we approximated the speaking rate of each video by dividing the total number of spoken words by the total in-video speaking time (i.e., words per minute).

Video type: We manually looked through each video and categorized its type as either an ordinary *lecture*, a *tutorial* (e.g., problem solving walkthrough), or *other* content such as a supplemental film clip. 89% of all videos were either lectures or tutorials, so we focus our analyses only on those two types.

Production style: We looked through each video and coded its production style using the following labels:

- **Slides** – PowerPoint slide presentation with voice-over
- **Code** – video screencast of the instructor writing code in a text editor, IDE, or command-line prompt
- **Khan-style** – full-screen video of an instructor drawing freehand on a digital tablet, which is a style popularized by Khan Academy videos
- **Classroom** – video captured from a live classroom lecture
- **Studio** – instructor recorded in a studio with no audience
- **Office Desk** – close-up shots of an instructor's head filmed at an office desk

Note that a video can contain multiple production styles, such as alternating between PowerPoint slides and an instructor's talking head recorded at an office desk. Thus, each video can have multiple labels.

Interviews With Domain Experts

To supplement our quantitative findings, we presented our data to domain experts at edX to solicit their feedback and interpretations. In particular, we conducted informal interviews with the four principal edX video producers who were responsible for overseeing all phases of video production—planning, filming, and editing. We also interviewed two program managers who were the liaisons between edX and the respective university course staff.

Public Anonymized Data Set and Scripts

We have uploaded an anonymized version of our data set along with analysis scripts and database installation instructions to **`http://www.pgbovine.net/edX/`** so that other researchers can reproduce and build upon this paper's findings. To our knowledge, ours is one of the first public data sets on MOOC resource usage.

FINDINGS AND RECOMMENDATIONS

We now detail the findings and recommendations of Table 1.

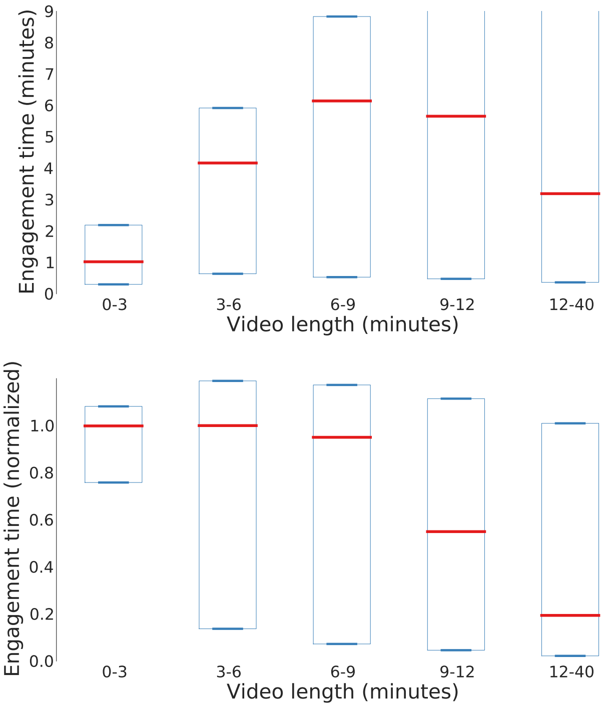

Figure 2. Boxplots of engagement times in minutes (top) and normalized to each video's length (bottom). In each box, the middle red bar is the median; the top and bottom blue bars are 25th and 75th percentiles, respectively. The median engagement time is at most 6 minutes.

Shorter Videos Are More Engaging

Video length was by far the most significant indicator of engagement. Figure 2 splits videos into five roughly equal-sized buckets by length and plots engagement times for 1x-speed sessions in each group[1]. The top boxplot (absolute engagement times) shows that median engagement time is at most 6 minutes, regardless of total video length. The bottom boxplot (engagement times normalized to video length) shows that students often make it less than halfway through videos longer than 9 minutes. The shortest videos (0–3 minutes) had the highest engagement and much less variance than all other groups: 75% of sessions lasted over three quarters of the video length. Note that normalized engagement can be greater than 1.0 if a student paused to check understanding or scrolled back to re-play an earlier portion before finishing the video.

To account for inter-courses differences, we made plots individually for the four courses and found identical trends.

Students also engaged less frequently with assessment problems that followed longer videos. For the five length buckets in Figure 2, we computed the percentage of video watching sessions followed by a problem attempt: The percentages were 56%, 48%, 43%, 41%, and 31%, respectively.

[1]Plotting all sessions pulls down the distributions due to students playing at 1.25x and 1.5x speeds and finishing videos faster, but trends remain identical. In this paper, we report results only for 1x-speed plays, which comprise 76% of all sessions. Our code and data are available to re-run on all sessions, though.

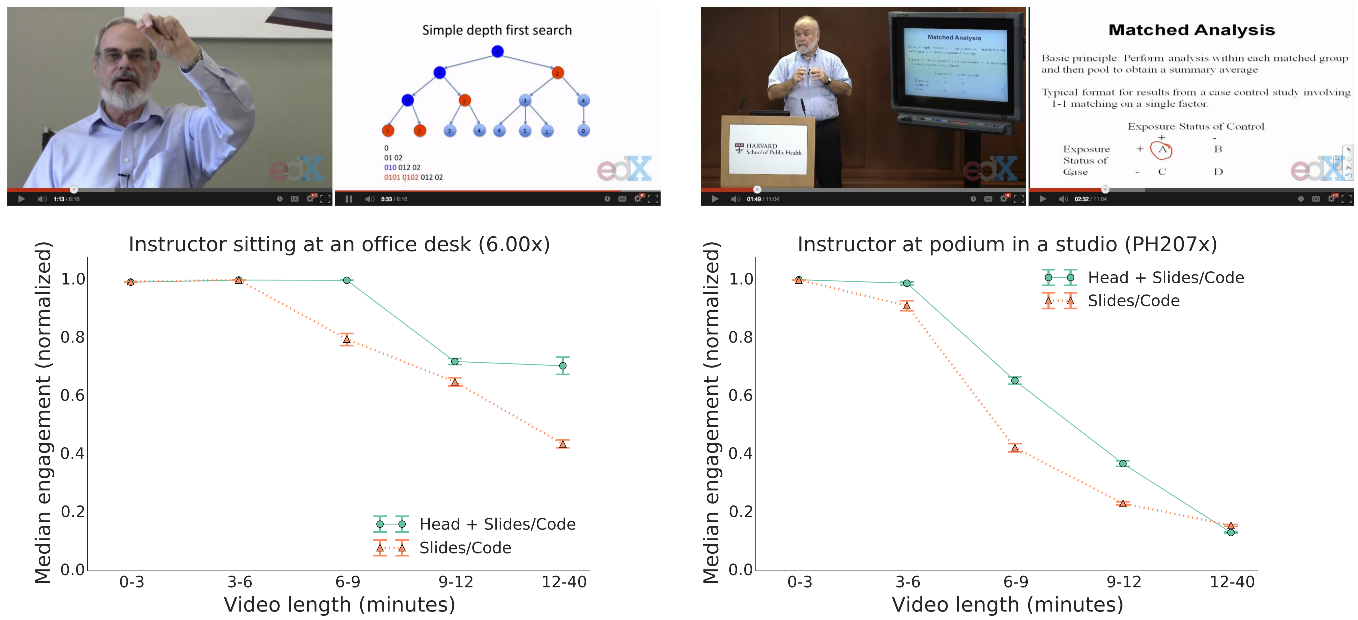

Figure 3. Median engagement times versus length for videos from 6.00x (left) and PH207x (right). In both courses, students engaged more with videos that alternated between the instructor's talking head and slides/code. Also, students engaged more with 6.00x videos, filmed with the instructor sitting at a desk, than with PH207x videos, filmed in a professional TV studio (the left graph has higher values than the right one, especially for videos longer than 6 minutes). Error bars are approximate 95% confidence intervals for the true median, computed using a standard non-parametric technique [14].

This particular set of findings resonated most strongly with video producers we interviewed at edX. Ever since edX formed, producers had been urging instructors to split up lessons into chunks of less than 6 minutes, based solely upon their prior intuitions. However, they often encountered resistance from instructors who were accustomed to delivering one-hour classroom lectures; for those instructors, even a 15-minute chunk seems short. Video producers are now using our data to make a more evidence-based case to instructors.

One hypothesis that came out in our interviews with video producers was that shorter videos might contain higher-quality instructional content. Their hunch is that it takes meticulous planning to explain a concept succinctly, so shorter videos are engaging not only due to length but also because they are better planned. However, we do not yet have the data to investigate this question.

For all subsequent analyses, we grouped videos by length, or else the effects of length usually overwhelmed the effects of other production factors.

> **Recommendation**: Instructors should segment videos into short chunks, ideally less than 6 minutes.

Talking Head Is More Engaging

The videos for two of our courses—6.00x and PH207x—were mostly PowerPoint slideshows and code screencasts. However, some of those videos (60% for 6.00x and 25% for PH207x) were edited to alternate between showing the instructor's talking head and the usual slides/code display.

Figure 3 shows that, in both courses, students usually engaged more with talking-head videos. In this figure and all subsequent figures that compare median engagement times, when the medians of two groups look far enough apart (i.e., their error bars are non-overlapping), then their underlying distributions are also significantly different ($p << 0.001$) according to a Mann-Whitney U test.

To check whether longer engagement times might be simply due to students pausing or re-playing the video, we compared the numbers of play/pause events in both groups and found no significant differences.

Also, 6.00x students attempted 46% of problems after watching a talking-head video (preceding a problem), versus 33% for other videos ($p << 0.001$ according to a chi-square test for independence). PH207x students attempted 33% of problems for both video groups, though.

These findings also resonated with edX video producers we interviewed, because they felt that a human face provided a more "intimate and personal" feel and broke up the monotony of PowerPoint slides and code screencasts. They also mentioned that their video editing was not done with any specific pedagogical "design patterns" in mind: They simply spliced in talking heads whenever the timing "felt right" in the video.

Since we have shown that this technique can improve engagement, we have encouraged producers to take a more systematic approach to this sort of editing in the future. Open questions include when and how often to switch between talking head shots and textual content. Perhaps video editing software could detect transition points and automatically splice in head shots. Finally, some people were concerned about the jarring effect of switching repeatedly between talking head and text, so a picture-in-picture view might work better.

Recommendation: Record the instructor's head and then insert into the presentation video at opportune times.

High Production Value Might Not Matter

Although 6.00x and PH207x were both taught by senior faculty at major research universities and had videos filmed in roughly the same style—slides/code with optional talking head—students engaged much more with 6.00x videos. The two graphs in Figure 3 show that students engaged for nearly twice as long on 6.00x videos between 6 and 12 minutes, and for nearly 3x the time on 6.00x videos longer than 12 minutes.

When we presented these findings to edX video producers and program managers who worked on those two courses, their immediate reaction was that differences in production value might have caused the disparities in student engagement: 6.00x was filmed informally with the instructor sitting at his office desk, while PH207x was filmed in a multi-million dollar TV production studio.

The "talking head" images at the top of Figure 3 show that the 6.00x instructor was filmed in a tight frame, often making direct eye contact with the student, while the PH207x instructor was standing behind a podium, often looking around the room and not directly at the camera. The edX production staff mentioned that the 6.00x instructor seemed more comfortable seated at his office having a personal one-on-one, office-hours style conversation with the video watcher. Video producers called this desirable trait "personalization"—the student feeling that the video is being directed right at them, rather than at an unnamed crowd. In contrast, the PH207x instructor looked farther removed from the watcher because he was lecturing from behind a podium in a TV studio.

The edX production staff worked with each instructor to find the recording style that made each most comfortable, and the PH207x instructor still preferred a traditional lecture format. Despite his decades of lecturing experience and comfort with the format, his performance did not end up looking engaging on video. This example reinforces the notion that what works well in a live classroom might not translate into online video, even with a high production value studio recording.

Here the supposed constraints of a lower-fidelity setting—a single close-up camera at a desk—actually led to more engaging videos. However, it is hard to generalize from only one pair of courses, since the effects could be due to differences in instructor skill. Ideally we would like to compare more pairs of low and high production value courses[2], but this was the only pair available in our data set.

Recommendation: Try filming in an informal setting where the instructor can make good eye contact, since it costs less and might be more effective than a professional studio.

Khan-Style Tutorials Are More Engaging

Now we focus on tutorials, which are step-by-step problem solving walkthroughs. Across all four courses, Khan-style tutorial videos (i.e., an instructor drawing on a digital tablet) were more engaging than PowerPoint slides and/or code screencasts. We group slides and code together since many tutorial videos feature both styles. Figure 4 shows that students engaged for 1.5x to 2x as long with Khan-style tutorials. For videos preceding problems, 40% of Khan-style tutorial watching sessions were followed by a problem attempt, versus 31% for other tutorials (chi-square $p << 0.001$). This finding corroborates prior work that shows how free-hand sketching facilitates more engaging dialogue [12] and how the natural motion of human handwriting can be more engaging than static computer-rendered fonts [4].

[2]or, even better, record one instructor using both styles.

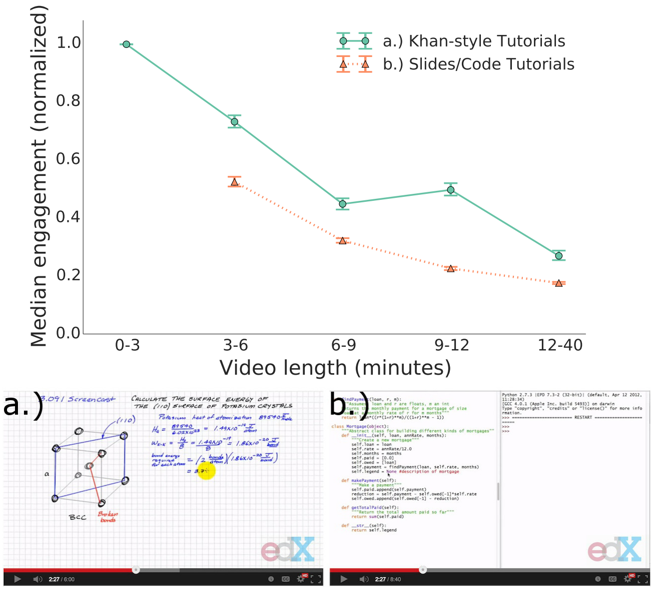

Figure 4. Median normalized engagement times vs. length for tutorial videos. Students engaged more with Khan-style tablet drawing tutorials (a.) than with PowerPoint slide and code screencast tutorials (b.). Error bars are approximate 95% confidence intervals for the true median [14].

Video producers and program managers at edX also agreed with this finding. In particular, they noticed how instructors who sketched Khan-style tutorials could situate themselves "on the same level" as the student rather than talking *at* the student in "lecturer mode." Also, one noted how a Khan-style tutorial "encourages professors to use the 'bar napkin' style of explanation rather than the less personal, more disjointed model that PowerPoint—if unintentionally—encourages."

However, Khan-style tutorials require more pre-production planning than presenting slides or typing code into a text editor. The most effective Khan-style tutorials were those made by instructors with clear handwriting, good drawing skills, and careful layout planning so as not to overcrowd the canvas. Future research directions include how to best structure Khan-style tutorials and how to design better authoring tools for creating and editing them. Perhaps some best practices from chalkboard lecturing could transfer to this format.

Recommendation: Record Khan-style tutorials when possible. If slides or code must be displayed, add emphasis by sketching over the slides and code using a digital tablet.

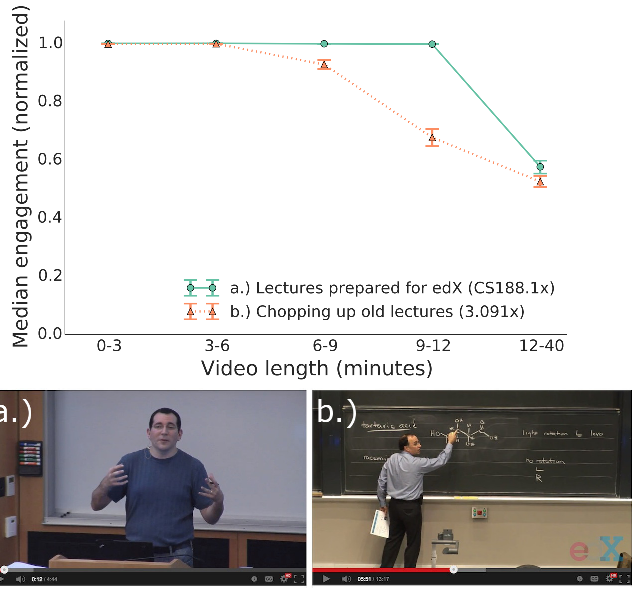

Figure 5. Median engagement times for lecture videos recorded in front of live classroom audiences. Students engaged more with lectures in CS188.1x (a.), which were prepared with edX usage in mind, than with lectures in 3.091x (b.), which were adapted from old lecture videos. Error bars are approximate 95% confidence intervals for true median [14].

Pre-Production Improves Engagement

So far, we have focused on production (i.e., filming) and post-production (i.e., editing) techniques that drive engagement. However, edX video producers we interviewed felt that the pre-production (i.e., planning) phase had the largest impact on the engagement of resulting videos. But since the output of extensive pre-production is simply better planned videos, producers cannot easily argue for its benefits by pointing out specific video features (e.g., adding motion via tablet sketches) to suggest as best practices for instructors.

To show the effects of pre-production, we compared video engagement for CS188.1x and 3.091x. Both are math/science courses with instructors who are regarded as excellent classroom lecturers at their respective universities. And both instructors wanted to record their edX lectures in front of a live classroom audience to bring out their enthusiasm. However, due to logistical issues, there was not enough time for the 3.091x instructor to record his lectures, so the video producers had to splice up an old set of lecture videos recorded for his on-campus class in Spring 2011. This contrast sets up a natural experiment where video recording styles are nearly identical, but no pre-production could be done for 3.091x.

Figure 5 shows that students engaged more with CS188.1x videos, especially longer ones. Also, for videos preceding problems, 55% of CS188.1x watching sessions were followed by a problem attempt, versus 41% for 3.091x (chi-square $p << 0.001$).

This finding resonated strongly with edX video producers, because they had always championed the value of planning lectures specially for an online video format rather than just chopping up existing classroom lecture recordings.

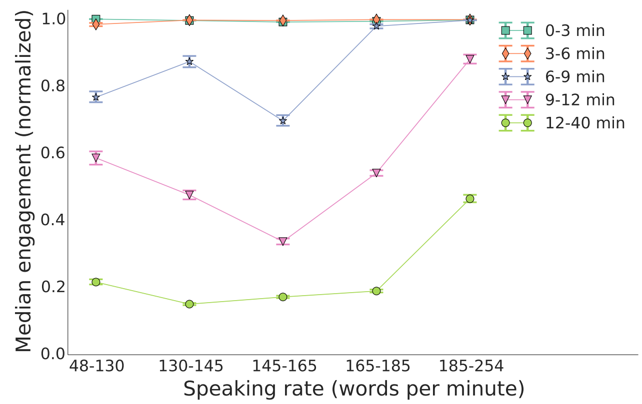

Figure 6. Median engagement times versus speaking rate and video length. Students engaged the most with fast-speaking instructors. Error bars are approximate 95% confidence intervals for the true median [14].

EdX staff who worked with the CS188.1x instructors reported that even though they recorded traditional one-hour lectures in front of a live classroom, the instructors carefully planned each hour as a series of short, discrete chunks that could easily be edited later for online distribution. In contrast, the 3.091x production staff needed to chop up pre-recorded one-hour lecture videos into short chunks, which was difficult since the original videos were not designed with the MOOC format in mind. There were often no clear demarcations between concepts, and sometimes material was presented out of order or interspersed with time- and location-specific remarks (e.g., "Jane covered this in last week's TA session in room 36-144") that broke the flow.

The main limitation here is that we had only one pair of courses to compare, and they differed in instructors and subject matter. To improve confidence in these findings, we could either find additional pairs to compare or, if the 3.091x instructor records new live lectures for edX, A/B test the engagement of old and new videos for that course.

> **Recommendation**: Invest in pre-production effort, even if instructors insist on recording live classroom lectures.

Speaking Rate Affects Engagement

Students generally engaged more with videos where instructors spoke faster. To produce Figure 6, we split videos into the usual five length buckets and also five equal-sized buckets (quintiles) by speaking rate. Speaking rates range from 48 to 254 words per minute (mean = 156 wpm, sd = 31 wpm). Each line represents the median engagement times for videos of a particular length range. As expected, students engaged less with longer videos (i.e., those lines are lower). Within a particular length range, engagement usually increases (up to 2x) with speaking rate. And for 6–12 minute videos, engagement dips in the middle bucket (145–165 wpm); slower-speaking videos are more engaging than mid-speed ones. Problem attempts also follow a similar trend, but are not graphed due to space constraints.

Some practitioners recommend 160 words per minute as the optimum speaking rate for presentations [15], but at least in our courses, faster-speaking instructors were even more engaging. One possible explanation is that the 160 wpm recommendation (first made in 1967) was for live lectures, but students watching online can actually follow along with much faster speaking rates.

The higher engagement for faster-speaking videos might also be due to students getting confused and re-playing parts. However, this is unlikely since we found no significant differences in the numbers of play and pause events among videos with different speaking rates.

To hypothesize possible explanations for the effects in Figure 6, we watched a random sample of videos in each speaking rate bucket. We noticed that fast-speaking instructors conveyed more energy and enthusiasm, which might have contributed to the higher engagement for those videos. We had no trouble understanding even the fastest-speaking videos (254 wpm), since the same information was also presented visually in PowerPoint slides. In contrast, instructors in the middle bucket (145–165 wpm) were the least energetic. For the slowest videos (48–130 wpm), the instructor was speaking slowly because he was simultaneously writing on the blackboard; the continuous writing motion might have contributed to higher engagement on those versus mid-speed videos.

Note that speaking rate is merely a surface feature that correlates with enthusiasm and thus engagement. Thus, speeding up an unenthusiastic instructor might not improve engagement. So our recommendation is not to force instructors to speak faster, but rather to bring out their enthusiasm and reassure them that there is no need to artificially slow down.

Video producers at edX mentioned that, whenever possible, they tightly edit in post-production to remove instances of "umm", "uhh", filler words, and other pauses, to make the speech more crisp. Their philosophy is that although speech pauses are beneficial in live lectures, they are unnecessary on video because students can always pause the video.

Recommendation: Work with instructors to bring out their natural enthusiasm, reassure them that speaking fast is okay, and edit out pauses and filler words in post-production.

Students Engage Differently With Lectures And Tutorials

Lecture videos usually present conceptual (declarative) knowledge, whereas tutorials present how-to (procedural) knowledge. Figure 7 shows that students only watch, on average, 2 to 3 minutes of each tutorial video, regardless of the video's length. Figure 8 shows that students re-watch tutorials more frequently than lectures.

These findings suggest that students will often re-watch and jump to relevant parts of longer tutorial videos. Adding hyperlink bookmarks or visual signposts on tutorial videos, such as big blocks of text to signify transitions, might facilitate skimming and re-watching. In contrast, students expect a lecture to be a continuous stream of information, so instructors should provide a good first-time watching experience.

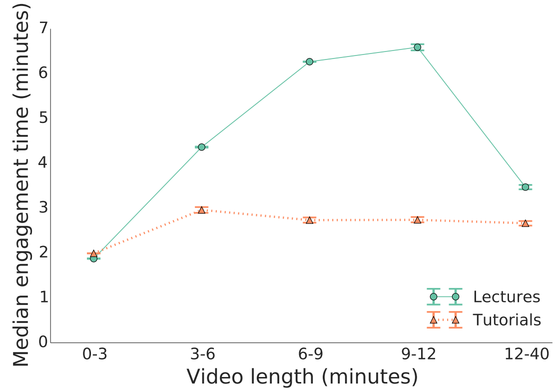

Figure 7. Median engagement times versus video length for lecture and tutorial videos. Students engaged with tutorials for only 2 to 3 minutes, regardless of video length, whereas lecture engagement rises and falls with length (similar to Figure 2). Error bars are approximate 95% confidence intervals for the true median [14].

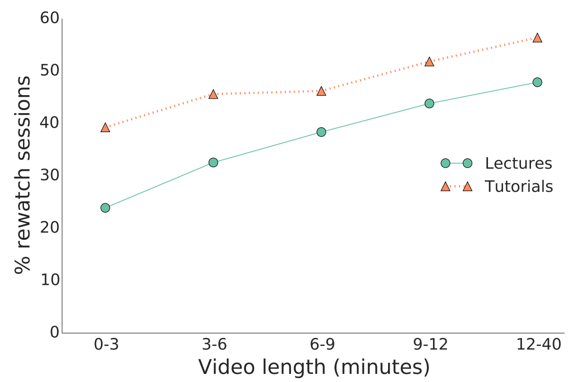

Figure 8. Percentage of re-watch sessions – i.e., not a student's first time watching a video. Tutorials were more frequently re-watched than lectures; and longer videos were more frequently re-watched. (Binomial proportion confidence intervals are so tiny that error bars are invisible.)

More generally, both our quantitative findings and interviews with edX staff indicate that instructors should adopt different production strategies for lectures and tutorials, since students use them in different ways.

Recommendation: For lecture videos, optimize the first-time watching experience. For tutorials, length does not matter as much, but support re-watching and skimming.

LIMITATIONS

This paper presents a retrospective study, not a controlled experiment. Also, we had access to the full server logs for only seven Fall 2012 edX courses, which were all math and science focused. Of those, we picked four courses with diverse production styles, subjects, and from different universities (Table 2). To improve external validity, these analyses should be replicated on additional, more diverse courses.

Our engagement findings might not generalize to all online video watchers, since edX students in the first Fall 2012 batch, who are more likely to be self-motivated learners and technology early adopters, might not be representative of the general online video watching population.

As we mentioned in the METHODOLOGY section, we cannot measure a student's true engagement with videos just from analyzing server logs. Our proxies—engagement time and problem attempts—might not be representative of true engagement. For instance, a student could be playing a video in the background while browsing Facebook. In the future, running a controlled lab study will provide richer qualitative insights about true engagement, albeit at small scale.

Also, we cannot track viewing activities of students who downloaded videos and watched offline. We know that the majority of students watched videos online in the edX video player, since the numbers in the "Students" column of Table 2 closely match the total enrollment numbers for each course. However, we do not have data on which students downloaded videos, and whether their behaviors differ from those who watched online.

Our data set contains only engagement data about entire videos. We have not yet studied engagement *within* videos such as which specific parts students are watching, skipping, or re-watching. However, we are starting to address this limitation in ONGOING WORK (see next section).

Lastly, it is important not to draw any conclusions about student learning solely from our findings about video engagement. MOOCs contain many components that impact learning, and different kinds of students value different ones. For instance, some learn more from discussion forums, others from videos, and yet others from reading external Web pages. The main relationship between video engagement and learning is that the former is often a prerequisite for the latter; if students are watching a video only for a short time, then they are unlikely to be learning much from it.

ONGOING WORK: WITHIN-VIDEO ENGAGEMENT

An alternative way to understand student engagement with MOOC videos is to measure how students interact with specific parts of the video. We have recently begun to quantify two dimensions of within-video interaction:

- **Interactivity** – How often do students pause the video while watching? To measure the degree of interactivity, we compute the mean number of pause events per second, per unique student. This metric controls for variations in viewer counts and video lengths. High interactivity could indicate more active engagement with the video content.
- **Selectivity** – Do students selectively pause more at specific parts of the video than others? This behavior might reflect uneven points of interest within the video. As a proxy for selectivity, we observe how the frequency of pause events vary in different parts of the video. Specifically, we compute the standard deviation of pause events across all seconds in a video. Higher selectivity videos attract more students to pause more at some parts than at others.

Here are two preliminary sets of findings. However, we have not yet interviewed edX production staff to get their interpretations or recommendations.

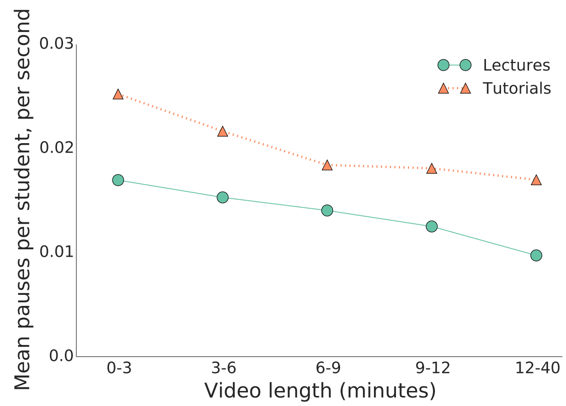

Figure 9. Students interacted (paused) more while watching tutorial videos than lecture videos.

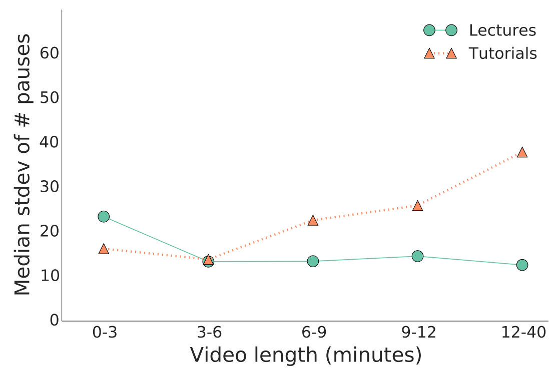

Figure 10. Students usually paused more selectively when watching tutorial videos than lecture videos.

Tutorial watching is more interactive and selective

Figure 9 shows that students interacted (paused) more within tutorial videos than lecture videos. This behavior might reflect the fact that tutorial videos contain discrete step-by-step instructions that students must follow, whereas lectures are often formatted as one continuous stream of content.

Figure 10 shows that students usually paused tutorial videos more selectively than lecture videos. This behavior might indicate that specific points in a tutorial video – possibly boundaries between distinct steps – are landmarks where students pause to reflect on or practice what they have just learned. This data could be used to automatically segment videos into meaningful chunks for faster skimming and re-watching.

Khan-style tutorials are more continuous

Figure 11 shows that students paused slides/code tutorials more selectively than Khan-style tutorials. One likely explanation is that Khan-style videos flow more continuously, so there are not as many discrete landmarks for pausing. In contrast, instructors of slides/code tutorials gradually build up text on a slide or a chunk of code, respectively, and then explain the full contents for a while before moving onto the next slide or code snippet; those are opportune times for pausing.

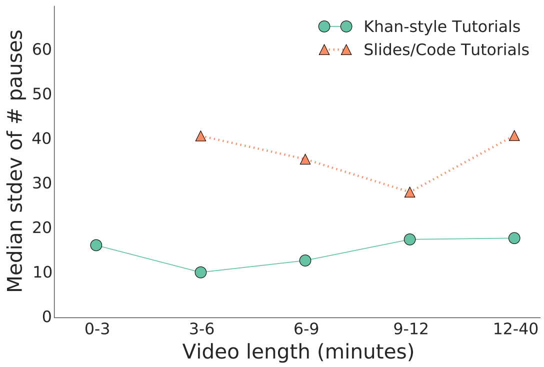

Figure 11. Students paused more selectively when watching slides/code tutorials than Khan-style tutorials.

Future Directions

Analyzing students' video interaction patterns allows educators to better understand what types of online videos encourage active interaction with content. The preliminary findings in this section provide an alternative perspective using micro-level, second-by-second interaction data that complements the engagement time analyses in the rest of this paper.

A possible future direction is to explore *why* students pause at certain points within the video. There are conflicting factors at play: Students might pause more because they consider a point to be important, or they might find the given explanation to be confusing and decide to re-watch until they understand it. Direct student observation in a lab setting could address these questions and complement our quantitative findings.

CONCLUSION

We have presented, to our knowledge, the largest-scale study of video engagement to date, using data from 6.9 million video watching sessions across four edX courses.

Our findings (Table 1) reflect the fact that, to maximize student engagement, instructors must plan their lessons specifically for an online video format. Presentation styles that have worked well for centuries in traditional in-person lectures do not necessarily make for effective online educational videos.

More generally, whenever a new communication medium arrives, people first tend to use it just like how they used existing media. For instance, many early television shows were simply radio broadcasts filmed on video, early digital textbooks were simply scanned versions of paper books, and the first online educational videos were videotaped in-person lectures. As time progresses, people eventually develop creative ways to take full advantage of the new medium. The findings from our study can help inform instructors and video producers on how to make the most of online videos for education.

Acknowledgments: Thanks to Anant Agarwal and our edX interview subjects for enabling this research, Olga Stroilova for helping with data collection, and Rob Miller for feedback.

REFERENCES

1. Khan Academy YouTube Channel. `http://www.youtube.com/user/khanacademy/about`.
2. Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. T. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research and Practice in Assessment 8* (Summer 2013).
3. Coetzee, D., Fox, A., Hearst, M. A., and Hartmann, B. Should Your MOOC Forum Use a Reputation System? CSCW '14, ACM (New York, NY, USA, 2014).
4. Cross, A., Bayyapunedi, M., Cutrell, E., Agarwal, A., and Thies, W. TypeRighting: Combining the Benefits of Handwriting and Typeface in Online Educational Videos. CHI '13, ACM (New York, NY, USA, 2013).
5. Google. How Visits are calculated in Analytics. `https://support.google.com/analytics/answer/2731565?hl=en`.
6. Google. YouTube Analytics. `http://www.youtube.com/yt/playbook/yt-analytics.html#details`.
7. Haber, J. xMOOC vs. cMOOC. `http://degreeoffreedom.org/xmooc-vs-cmooc/`, 2013.
8. Ilioudi, C., Giannakos, M. N., and Chorianopoulos, K. Investigating Differences among the Commonly Used Video Lecture Styles. In *Proceedings of the Workshop on Analytics on Video-based Learning*, WAVe '13 (2013).
9. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, ACM (New York, NY, USA, 2013), 170–179.
10. Levasseur, D. G., and Sawyer, J. K. Pedagogy Meets PowerPoint: A Research Review of the Effects of Computer-Generated Slides in the Classroom. *Review of Communication 6*, 1 (2006), 101–123.
11. Mayer, R. E. *Multimedia Learning*. Cambridge University Press, 2001.
12. Roam, D. *The Back of the Napkin (Expanded Edition): Solving Problems and Selling Ideas with Pictures*. Portfolio Hardcover, 2009.
13. Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., and Pritchard, D. E. Who does what in a massive open online course? *Communications of the ACM* (2013).
14. Wade, A., and Koutoumanou, E. Non-parametric tests: Confidence intervals for a single median. `https://epilab.ich.ucl.ac.uk/coursematerial/statistics/non_parametric/confidence_interval.html`.
15. Williams, J. R. Guidelines for the use of multimedia in instruction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 42*, 20 (1998), 1447–1451.
16. Wistia. Does length matter? It does for video! `http://wistia.com/blog/does-length-matter-it-does-for-video`, Sept. 2013.