122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Seattle Making Value for Society
Paper ID #13527

The Myth of the Six Minute Rule: Student Engagement with Online Videos

Dr. Larry Lagerstrom, Stanford University

Larry Lagerstrom is the Director of Online Learning for the School of Engineering at Stanford University. Before coming to Stanford he taught computer programming and electrical engineering for sixteen years at U.C. Berkeley and U.C. Davis. He has degrees in physics, math, history, and interdisciplinary studies, including a PhD in the history of science and technology. He also recently taught a MOOC on "Understanding Einstein: The Special Theory of Relativity."

Petr Johanes, Stanford University
Mr. Umnouy Ponsukcharoen

The Myth of the Six-Minute Rule: Student Engagement with Online Videos

Introduction

In the era of massive open online courses (MOOCs), video has become a standard medium for production, instruction, and experimentation in online and blended education. The employment of video in distance education is hardly new, of course, but researchers are now taking advantage of the influx of MOOC data to look more closely at video design, use, and efficacy. Videos are a critical resource for students in online learning environments. A breakdown of student activity from edX's first MOOC revealed that certificate earners spent most of their weekly time watching lecture videos and working on homework assignments. Lecture videos also served as one of the most frequently accessed resources for students working on the course homework assignments.[1]

To study interactions with videos, researchers have designed specialized interfaces that map the interaction history of students.[4, 5, 7, 9] One study, for example, focused specifically on video interaction through peaks in user events and analyzing the peak profile through the peak height, width, and area. This analysis revealed five peak types, from which the authors provide recommendations for content authoring and interface design.[3] Other studies have focused on how to use these data to inform student and instructor behavior, focusing on feeding back data through a dashboard,[2] multimedia exercises,[5] and three-dimensional video cubes.[11]

Some of these studies attempt to gauge student attention and engagement or correlate them with other metrics. A common rule of thumb for instructors producing videos has been that when it comes to capturing and retaining students' attention, shorter is better. This advice is partly based on the popular culture of web videos, with data showing a marked drop-off in viewership after several minutes of a video.[13] It is also based on the so-called "ten-minute rule": during a presentation audience members tend to lose focus after about ten minutes, unless re-stimulated by a compelling hook.[10] A recent study has refined the ten-minute rule into a six-minute rule. Based on an analysis of 6.9 million MOOC video viewing episodes, the authors found that "median engagement time is at most six minutes, regardless of total video length." They recommend that instructors "should segment videos into short chunks, ideally less than 6 minutes."[2]

We call this "the myth of the six-minute rule" not because it is false. The evidence and analysis show that most MOOC videos do not retain students' attention beyond about six minutes. Rather, we are using "myth" in the sense of a powerful idea that gets at truth, but that can become more false than true if it hardens into a universal principle for all times and circumstances. The nice sound-bite quality of "the six-minute rule" (or "however-many-minute rule") makes it susceptible to turning into a viral meme that thereby becomes an actual rule. Almost immediately after the MOOC video study was published and received media coverage, we heard the six-minute recommendation being cited by a university leader. Repeated enough times by enough influential individuals, it becomes a tacit norm.

Restricting instructional videos to six to ten minutes in length make sense from multiple perspectives, including production (less logistical difficulty), quality (more concision/clarity), and psychology (lower attention/persistence demands). And yet audiences maintain attention through eighteen-minute TED talks, two-hour movie showings, and, yes, sometimes even 50-minute university lectures. In addition, notwithstanding the impressive dataset underlying the six-minute recommendation (6.9 million videos), the diverse nature of the MOOC audience is an important factor. People enroll in MOOCs for a variety of reasons. Some are seeking a serious instructional experience, while others may be interested in the MOOC as a free refresher course on previously studied material, and still others may intend only to sample the covered topics here and there. Differing intentions and expectations will lead to differing video viewing patterns. The question then arises as to the nature of viewing patterns in instructional settings where students have more homogeneous intentions, such as a for-credit course in which all students have paid tuition or a fee to enroll. This type of instructional setting corresponds more closely than the MOOC setting to the usual higher education setting. We therefore undertook to explore the question of student engagement and optimum video length in a context more typical of higher education.

Methodology

For this study, we focused on two computer science courses offered during the Autumn 2014 quarter at Stanford University: an introductory level course ("CS100") and a more advanced course ("CS200"), each with enrollments of several hundred students. The students included regular undergraduate and graduate students, as well as working professionals who are allowed to enroll in select engineering courses via the Stanford Center for Professional Development (SCPD). The SCPD professionally videotapes and edits the course lectures and makes them available to all of the course's students via its website. Most of the videos for the CS100 course were around 50 minutes long, and most of the videos for the CS200 course were around 75 minutes long. Watching the videos was not required, as all students were welcome (though not required) to attend the live lectures. Nevertheless, many students chose to watch the videos, either in lieu of attending the lectures or as a supplement to them. And, of course, the videos were the only access to the lectures for most if not all of the working professionals. Both courses had multiple teaching assistants and well-established support mechanisms for students. Assignments included coding projects and homework assignments, as well as mathematically-driven problem sets for the advanced course.

For CS100, there were 28 videos, with 13,107 watching sessions by the students. For CS200, there were 29 videos and 17,034 watching sessions. Following Reference 2, a "watching session" is defined as a single instance of a student watching a particular video. The start of the session is defined when the student initiates a "play" event. The end of the session may occur in one of several ways, such as when the end of the video is reached, the student navigates away from the page, or the student ends the login session. In addition, if a 30-minute gap occurs between video events (e.g., a 30-minute pause), it is assumed that the user is no longer engaged with the video and the session is ended. (As explained below, the video player has a "heartbeat" minute marker, which marks each minute passed in the video while it is playing. While playing, therefore, video events occur at least once per minute. Of course, it's possible that a student would start a video and leave it running while doing something else. We assume that this does not happen often.)

Both courses used the OpenEdX platform combined with a custom-built video player created by the SCPD platform team for data collection purposes. We worked closely with the SCPD team to test and improve the data collection capabilities of the player. The platform team provided video log files that included various pieces of data that generally fall into either a basic or an interaction category. In the basic category, the log files included video identification markers, log-in session markers, hashed user names, and other data that allowed us to precisely isolate the user, video, and log-in session. In the interaction category, the log files included the nature of the user's interaction with the video, real-world timestamp, video timestamp, and other data that allow us to define how the user interacted with the video. The interactions we tracked were:

- Load – Whenever the web browser loads the video, the database records this as a load. The recorded value is binary (0 or 1).
- Pause – Whenever the user clicks on the pause button, the database records this as a pause. The recorded value is the video timestamp.
- Play – Whenever the user clicks on the play button, the database records this as a play. The recorded value is the video timestamp.
- Resolution change – Whenever the user clicks on the resolution change button, the database records this as a resolution change. The recorded value is the video timestamp and the resolution the user changed to.
- Speed change – Whenever the user clicks on the speed change button, the database records this as a speed change. The recorded value is the video timestamp and the speed the user changed to.
- End – Whenever the video finishes (i.e., reaches the end of the video timestamp), the database records it as an end to the video. The recorded value is binary (0 or 1).
- Jump – Whenever the user clicks on a location in the video player bar, the database records this as a jump. The recorded value is the video timestamp for where the user is jumping from (i.e., the video time at the moment of the click in real time)
- Land – Whenever the user clicks on a location in the video player bar and the video skips to that location, the database records this as a land. The recorded value is the video timestamp for where the user is jumping to (i.e., the video time at the location of the mouse cursor).
- Minute Marker – Whenever the video player bar indicator passes a full minute (at 60 seconds, 120 seconds, and so on), the database records this as passing a minute marker. The recorded value is the minute value of the marker (e.g. 1 for passing 60 seconds, 2 for passing 120 seconds, etc.).

An initial analysis uncovered anomalies in approximately 30 percent of the video viewing data. We therefore put a process in place to identify, categorize, replicate, and correct them. The anomalies fell into one of the following three categories:

- Signal Delay – This anomaly manifests itself in the form of a discrepancy between the real and the video time durations of an event. A classic example in this category is that the database records the user watching 60 seconds of video while 75 seconds pass in real time. This influences subsequent recorded actions in that all of them are offset by the same amount of time (by 15 seconds in the example). We attribute this kind of anomaly to buffering and/or other Internet bandwidth issues.
- Concurrent Events – This anomaly manifests itself in the form of multiple events being recorded at the same video time. A typical example in this category is that the database records the user passing minute markers for minutes 3, 4, 5, and 6 at the 6-minute marker. What usually happens is that the real times associated with passing each of those minute markers are correct (i.e. not at the same point in time, but spaced out appropriately). Therefore, we attribute the anomaly to a temporary overload of the database recording system, when the database has to catch up due to too many calling requests for data.
- Illogical Order – This anomaly manifests itself in the form of a sequence of events being recorded that does not logically follow the recording rules. One example in this category is seeing a jump action and then a play action without a land action or a pause action in between them. Another example would be observing a land action at minute 6 followed by a pause action at minute 2. The illogical order, therefore, can be as much in the sequencing as the timing of the actions. This particular kind of anomaly took us longer to replicate and understand, and we were eventually able to attribute it to scrubbing. Scrubbing is when the user clicks and holds the video player bar indicator and slides it to another point in the video. This is different from a jump-and-land action in that this is not an instantaneous click event in a single location, but, rather, an extended click event that takes the mouse cursor from one location to another.

After identifying the causes of the anomalies, we worked with the SCPD platform team to understand if the anomalies represented data that were mis-recorded or mis-collected in the first place. If they were merely mis-recorded in the database, then they could be cleaned up in later analysis, but if they were mis-collected in the video player, then our ability to interpret them would be limited. To that end, the video player development team created a separate sandbox version of the video player and database. This version allowed us to play a video and interact with it in one window and see our actions show up in the database in another window. We used this sandbox version to reproduce the anomalies we found in our data to confirm the conclusions listed above. The watching sessions mentioned above and on which our analysis is based (13,107 for CS100 and 17,034 for CS200) were the clean sessions that resulted after the anomalies were eliminated.

Results and Discussion

We analyzed the video viewing data at both the user and video levels. At the user level, we graphed the sequences of actions each individual user took in a particular video. Figures 1(a)-(c) below showcase plots of some of the possible user behaviors: multiple cases of pausing and playing, multiple cases of speed changes, and multiple cases of jumping. This analysis helped us uncover anomalies in the data (listed above) as well as patterns of interactions. Creating visualizations of these actions for each user also proved a useful exercise in testing different ways of presenting this type of information for an instructor or researcher dashboard, for instance.

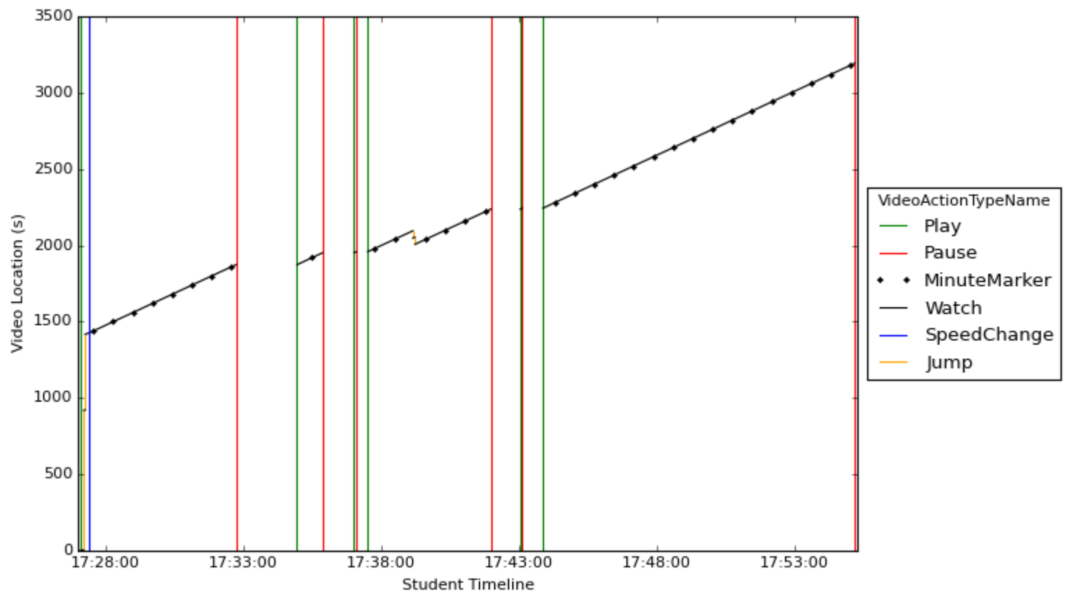

Figure 1(a). Visualization of a single user's behavior during a video watching session, showing multiple cases of pausing (vertical red lines) and playing (vertical green lines). The passing of each minute marker as the student watches the video is represented by a black dot. Horizontal gaps, such as the gap after the pause around 17:33:00, represent longer pauses. (Video Location on the vertical axis represents the user's location within the video in seconds, while Student Timeline on the horizontal axis represents real time in hours:minutes:seconds.)

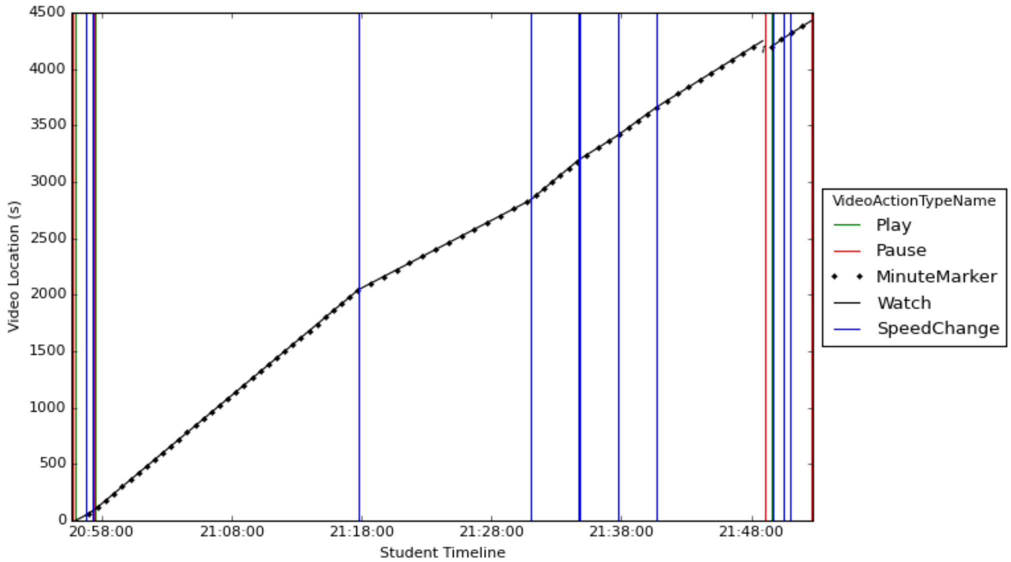

Figure 1(b). Visualization of a single user's behavior during a video watching session, showing cases of speed changing (vertical blue lines mark the time of each speed change). The passing of each minute marker as the student watches the video is represented by a black dot. (Video Location on the vertical axis represents the user's location within the video in seconds, while Student Timeline on the horizontal axis represents real time in hours:minutes:seconds.)

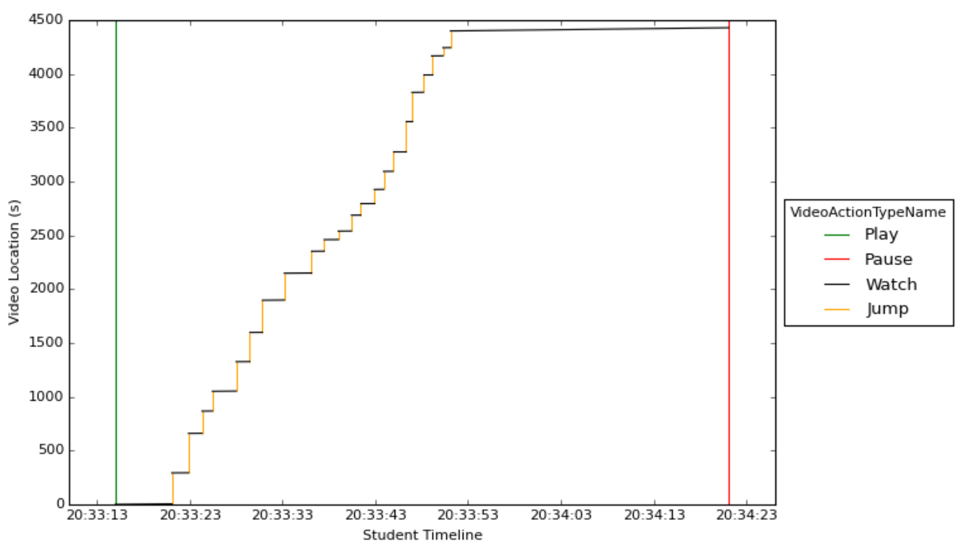

Figure 1(c). Visualization of a single user's behavior during a video watching session, showing multiple cases of jumping. Note that the black "watching" lines seem horizontal due to the relatively short timescale on the horizontal axis in this example (the whole horizontal axis represents an elapsed time of just over one minute); the lines actually slope slightly upward. (Video Location on the vertical axis represents the user's location within the video in seconds, while Student Timeline on the horizontal axis represents real time in hours:minutes:seconds.)

To explore patterns of student interaction with individual videos, we aggregated the actions of the users who watched a given video. Figure 2(a) shows one way to visualize student engagement with the video. The red curve represents the number of unique users who watched a given point in the video (count of users on the vertical axis vs. the video timeline in seconds on the horizontal axis). In this example we see that approximately 220 students started watching the video (0 point on timeline), which quickly fell to just below 200 students after the opening seconds. It then held more or less steady until the end of the video at the 4650-seconds mark. In other words, approximately 90 percent of the students who started the video watched all or nearly all of its 77 minutes (though not necessarily without pauses, jumps, or multiple sessions of watching to get through the whole video).

The blue curve in Figure 2(a) represents the total number of watches for each point in the video's timeline. Peaks in the blue curve therefore indicate sections of the video that were watched repeatedly. Such information can be useful to an instructor by potentially indicating concepts that gave the students trouble and/or a presentation that lacked clarity.

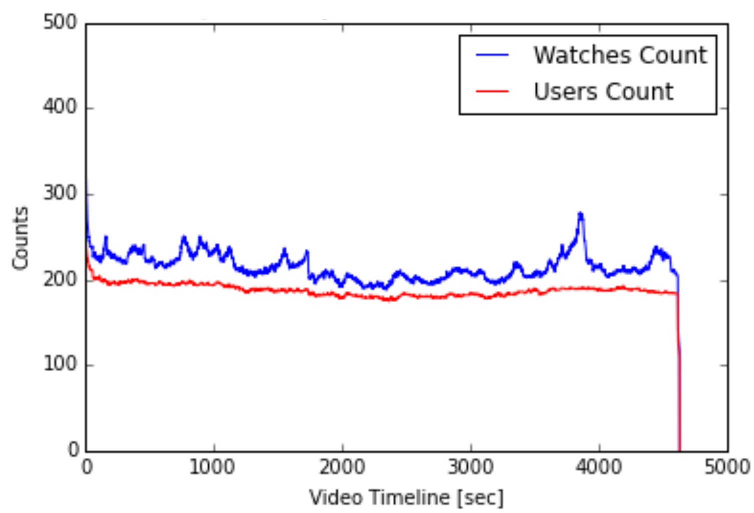

Figure 2(a). Graph showing aggregated user interactions with a given video. The red (bottom) curve tracks views by unique users and the blue (top) curve tracks the total number of watchings (i.e., it tallies multiple views by individual users). The horizontal axis is the video timeline in seconds and the vertical axis is the count of views.

Figure 2(b) shows a similar plot for a different video. In this case approximately 150 students started watching the video, but the viewership had dropped by almost two-thirds by the 10-minute (600 seconds) mark. It then jumps up again at around the 30-minute (1800 seconds) mark. Clearly many students skipped over the sections of the video from approximately 10 minutes to 30 minutes. In addition, the close tracking of the red curve and blue curve indicates that few students repeated sections of the video, until the last few minutes when several blue peaks push above the red curve.

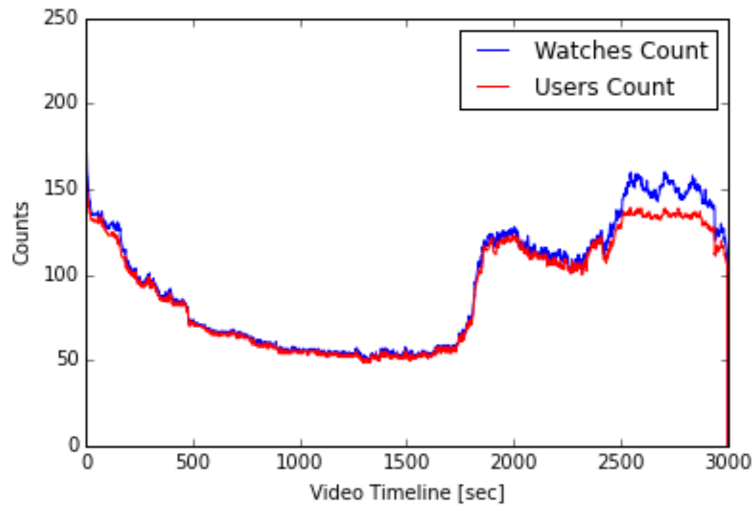

Figure 2(b). Graph showing aggregated user interactions with a given video (different video than that shown in Figure 2(a)). The red curve tracks views by unique users and the blue curve tracks the total number of watchings (i.e., it tallies multiple views by individual users). The horizontal axis is the video timeline in seconds and the vertical axis is the count of views.

Figure 3 represents another way to gain insight into student interaction with a given video. It uses a heat map to answer the question: When students choose to jump from one place to another in a video, where do they most often land? The horizontal axis plots the video time (in minutes) at which the student chose to jump (either forwards or backwards), and the vertical axis plots the video time at which the student lands. So, for example, a hot spot at the point (30, 25) would indicate that a number of students jumped from the 30-minute mark back to the 25-minute mark.

The sample plot in Figure 3 shows a heat map with a fairly well-defined line of slope 1. Such a result indicates that for this video most of the jumps, whether backwards or forwards, were short (i.e., the jumping off point was almost equal to the landing point). Points above the slope-1 line represent forward jumps (landing time is greater than jumping off time) and points below the slope-1 line represent backward jumps (landing time is less than jumping off time). Darker points indicate video segments with large numbers of jump starts and/or jump landings, and therefore potentially indicate video sections where students are confused or having trouble. A hot spot above the slope-1 line would likely indicate a section that students believe is particularly important or interesting, i.e., many students jump ahead to get to that point. Conversely, a hot spot below the slope-1 line would indicate that later in the video many students realized that a previous section was crucial for understanding, and therefore they jumped back to it.

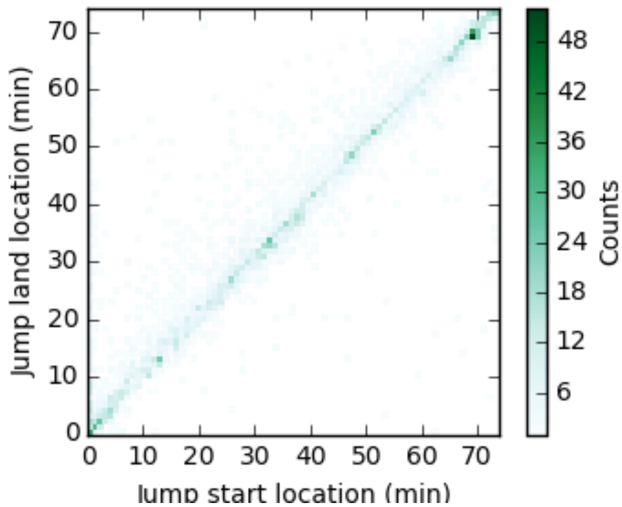

Figure 3. Heat map graph that aggregates user jumps for a given video. The horizontal axis plots the video time (in minutes) at which the student chose to jump (either forwards or backwards), and the vertical axis plots the video time at which the student lands.

Moving from aggregated student-viewing data for individual videos to aggregated data for all the videos in CS100 and CS200, what do the results tell us about the six-minute rule?

In the 13,107 sessions from the 28 CS100 videos, we find a median watching time of 12.1 minutes, i.e., the number of video minutes viewed during a watching session. In addition, the median total time for a watching session (i.e., actual time spent, including pauses and other actions) was 16.1 minutes. The mean watching time was 17.4 minutes and the mean total time spent on a session was 22.4 minutes. (The average video length was 50.1 minutes.)

In the 17,034 watching sessions from the 29 CS200 videos, the median watching time was 13.0 minutes and the median total time spent on a session was 19.5 minutes. The mean watching time was 19.5 minutes and the mean total time spent on a session was 26.9 minutes. (The average video length was 66.4 minutes.)

In the previously cited study of MOOC video watching behavior,[2] the researchers reported that engagement time decreases with increasing video length and that in the best case scenario, involving videos of length 6-9 minutes, student engagement tends to peak at around 6 minutes. For videos 12-40 minutes long, they found a median engagement time of only 3 minutes. The videos in our study were longer—most either 50 or 75 minutes—and we have found median engagement times that are approximately four times higher.

But the use of "watching sessions" here hides an even more significant result. As discussed earlier, a watching session represents a single session of a student's interaction with a video. Given that the median watching time for a session was 12-13 minutes, and most of the videos were 50-75 minutes long, clearly many students were not watching a whole video in one session. It raises the question, however, of how many students actually watched most or all of a given video. Does the 12-13-minute median result mean that many students gave up and never got around to watching the rest of the video? In fact, not. Most students had multiple watching sessions with each individual video, and when those multiple sessions are stitched together, a truer picture emerges. In the discussion of Figure 2(a) above, we noted that it represented a video where approximately 90 percent of the students watched nearly the whole video. This turned out to be a common result for the majority of the videos in our study. We found that for most videos, most students watched 90-100 percent of the video. In addition, we found that the students who did not fall in this category watched very little of the video, usually less than 10 percent. In other words, there is a bimodal distribution with most students either watching the video in its entirety (usually over multiple sessions) or not at all (Figure 4).

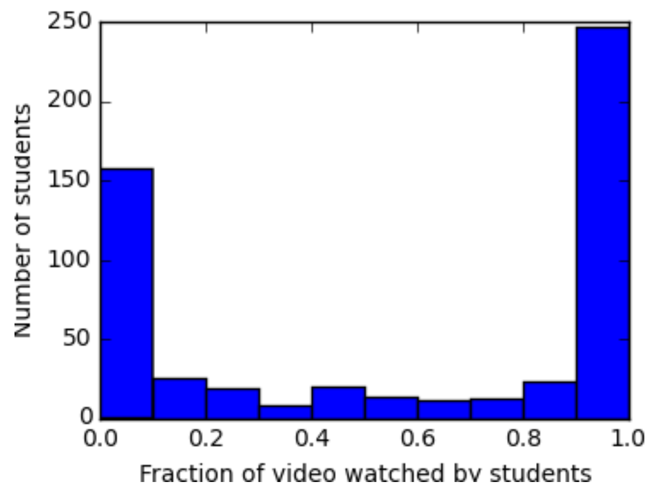

Figure 4. Shows the dominant trend regarding how much of a video students watch: For the represented video (and for most videos), the majority of students watched 90-100 percent of it. Most of the rest of the students watched less than 10 percent of it.

It is therefore instructive to divide the student population in each class (CS100 and CS200) into deciles based on the fraction of a given video that students watch. Tables 1 and 2 below provide the results for the 28 videos of CS100 and the 29 videos of CS200, respectively. The 0-10% column records the percentage of students who watched up to 10% of the video, the 10-20% column records the number who watched from 10.1% up to 20%, and so on. (The video watching is not necessarily linear, of course, even within a single watching session. Many students periodically paused and/or jumped while watching a video.)

The key columns in both tables are the “90-100%” and “Median % Viewed” columns. They show that, with a few exceptions, a majority of the students ended up watching 90-100% of each video, and that the median of the percentage viewed for most of the videos was in the high 90s. For CS100 (Table 1), the mean of the 28 “Median % Viewed” values is 91.3%. For CS200 (Table 2), the mean of the 29 “Median % Viewed” values is 74.0% (a few very low values pull this mean down).

Video	Length (mins)	0-10%	10-20%	…	80-90%	90-100%	Median Total % Viewed	Mean Total % Viewed	Median Minutes Viewed	Mean Minutes Viewed
1	51.0	45.3	8.6	…	0.8	17.2	15.6	31.7	7.9	16.2
2	51.2	24.4	8.1	…	0.8	54.5	93.9	62.2	48.1	31.8
3	50.4	21.5	3.7	…	3.0	62.2	97.8	69.7	49.3	35.1
4	50.2	14.6	2.5	…	3.8	54.4	97.4	70.8	48.9	35.5
5	51.7	12.4	4.7	…	4.7	65.7	98.0	75.7	50.7	39.1
6	51.2	9.8	2.8	…	5.6	68.4	98.6	79.6	50.4	40.7
7	51.6	15.3	4.9	…	1.5	60.6	98.0	71.4	50.5	36.8
8	50.2	9.7	3.2	…	3.2	66.4	98.5	78.4	49.4	39.3
9	50.6	10.5	3.4	…	3.7	71.3	98.8	80.0	50.0	40.4
10	51.0	12.9	4.0	…	6.8	63.9	97.6	76.3	49.7	38.9
11	51.3	13.6	5.8	…	4.3	56.2	95.8	71.2	49.2	36.5
12	50.2	12.5	1.1	…	6.0	68.7	98.0	79.7	49.2	40.0
13	49.4	8.8	3.2	…	3.9	71.4	99.1	81.1	49.0	40.1
14	49.3	8.7	2.7	…	5.3	70.1	98.6	81.9	48.6	40.3
15	47.2	11.1	3.1	…	6.6	60.8	97.5	74.9	46.0	35.3
16	49.5	8.6	1.9	…	15.6	55.8	92.9	78.2	46.0	38.7
17	55.8	10.3	1.3	…	35.8	38.7	89.0	75.2	49.6	41.9
18	50.6	14.5	3.3	…	11.6	49.6	89.8	71.2	45.4	36.0
19	50.8	9.6	4.1	…	6.9	68.0	97.2	80.3	49.4	40.7
20	47.8	9.8	1.0	…	7.4	71.4	98.3	83.5	47.0	39.9
21	48.2	9.6	4.1	…	4.8	69.5	98.6	79.8	47.5	38.5
22	49.6	8.5	3.1	…	5.4	66.7	97.5	79.4	48.4	39.3
23	48.8	7.4	3.5	…	4.8	71.9	98.4	82.6	48.0	40.3
24	48.3	16.2	2.7	…	4.2	67.3	98.3	76.3	47.5	36.8
25	50.5	8.7	8.7	…	4.0	65.0	98.7	76.2	49.9	38.5
26	49.2	9.6	3.1	…	9.6	55.4	92.8	75.4	45.7	37.1
27	46.6	15.8	12.9	…	4.1	46.8	81.8	62.2	38.1	29.0
28	50.0	17.4	13.2	…	2.4	22.8	40.1	46.5	20.0	23.2
Ave.	50.1	13.5	4.5	…	6.3	59.3	91.3	73.3	45.7	36.7

Table 1. Results for the total fraction of video watched by the students in CS100, divided into deciles (28 videos). The values in columns 3, 4, 6, and 7 represent the percentage of students who watched the indicated percentage of the video. For example, the values in the 0-10% column represent the percentage of students who watched only 0-10% of the video, while the values in the 90-100% column represent the percentage of students who watched 90-100% of the video. The 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, and 70-80% columns are hidden for visualization and formatting purposes. Most of the values in the hidden columns are on the order of 1-3 percent. The final four columns give the median and mean percentage of each video viewed by the students, as well as the median and mean number of minutes viewed.

Video	Length (mins)	0-10%	10-20%	…	80-90%	90-100%	Median Total % Viewed	Mean Total % Viewed	Median Minutes Viewed	Mean Minutes Viewed
1	73.8	29.3	4.7	…	4.3	46.2	80.9	58.0	59.7	42.8
2	72.6	19.2	3.9	…	2.9	64.9	98.5	72.0	71.5	52.3
3	51.5	23.4	6.2	…	2.0	47.5	76.6	59.8	39.4	30.8
4	76.2	18.6	3.5	…	2.6	65.2	98.7	72.5	75.2	55.2
5	74.7	18.2	1.8	…	2.9	64.4	98.4	73.6	73.5	55.0
6	50.9	41.2	5.2	…	0.5	38.9	29.9	46.9	15.2	23.9
7	75.4	12.3	4.5	…	3.6	70.5	98.7	78.4	74.4	59.1
8	75.3	17.7	2.2	…	3.6	63.3	98.6	72.8	74.2	54.8
9	53.4	43.9	5.7	…	3.8	28.3	21.0	41.7	11.2	22.3
10	76.8	13.1	0.3	…	3.8	74.1	99.0	80.9	76.0	62.1
11	73.8	15.9	1.4	…	3.1	69.0	98.9	77.4	73.0	57.1
12	39.5	26.1	5.9	…	3.7	47.3	86.9	59.5	34.3	23.5
13	76.5	18.8	3.5	…	2.1	63.0	98.4	70.9	75.3	54.2
14	76.4	20.2	1.9	…	2.8	67.0	98.7	73.5	75.3	56.1
15	40.7	39.6	4.5	…	0.0	41.8	47.8	49.7	19.5	20.2
16	77.2	17.2	2.1	…	3.6	67.7	98.7	75.8	76.2	58.5
17	75.0	21.5	3.9	…	3.0	54.9	97.0	65.6	72.8	49.2
18	50.9	26.8	3.9	…	4.4	33.2	47.4	51.7	24.1	26.3
19	71.8	27.4	3.4	…	1.9	53.9	95.2	62.4	68.4	44.8
20	77.1	22.0	1.9	…	6.7	58.1	97.0	69.8	74.8	53.8
21	51.5	38.3	2.5	…	1.2	51.9	96.0	56.7	49.5	29.2
22	77.2	27.3	5.3	…	3.9	48.9	86.6	60.9	66.8	47.0
23	77.2	28.1	2.2	…	0.9	58.0	97.8	64.4	75.6	49.7
24	42.0	58.5	0.0	…	2.4	30.5	2.5	38.1	1.0	16.0
25	75.5	19.6	2.8	…	4.2	57.5	96.7	69.4	73.0	52.4
26	73.9	27.3	2.0	…	4.5	47.5	84.8	60.8	62.7	44.9
27	43.4	69.0	8.0	…	1.1	13.8	3.2	21.1	1.4	9.2
28	75.8	64.8	5.6	…	0.9	19.4	3.7	26.5	2.8	20.1
29	70.9	56.4	10.3	…	0.0	20.5	7.9	27.7	5.6	19.6
Ave.	66.4	29.7	3.8	…	2.8	50.6	74.0	59.9	51.8	41.0

Table 2. Results for the total fraction of video watched by the students in CS200, divided into deciles (29 videos). The values in columns 3, 4, 6, and 7 represent the percentage of students who watched the indicated percentage of the video. For example, the values in the 0-10% column represent the percentage of students who watched only 0-10% of the video, while the values in the 90-100% column represent the percentage of students who watched 90-100% of the video. The 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, and 70-80% columns are hidden for visualization and formatting purposes. Most of the values in the hidden columns are on the order of 1-3 percent. The final four columns give the median and mean percentage of each video viewed by the students, as well as the median and mean number of minutes viewed.

The results tabulated in Tables 1 and 2 represent all the enrolled students in CS100 and CS200, respectively. As previously noted, the population of students in these classes consists of both regular university students and working professionals. We might expect that because the working professionals do not have easy access to the live course lectures, their results would show an even stronger preference for watching 90-100% of any given video. (Nearly all undergraduate students live in on-campus housing, and most graduate students live in on-campus or nearby housing, so that access to campus is relatively easy for them.)

To test this assumption, we identified the two groups of students in the data based on their different login status and then looked at how much of the CS200 videos each set of learners watched. Figure 5 shows the results. On average, a higher percentage of the working professionals watched 50% or more of the total length of each video, compared to the regular university students. At the same time, the distributions are not that different, both sporting high frequencies at each end (0-10% and 90-100% of video watched).

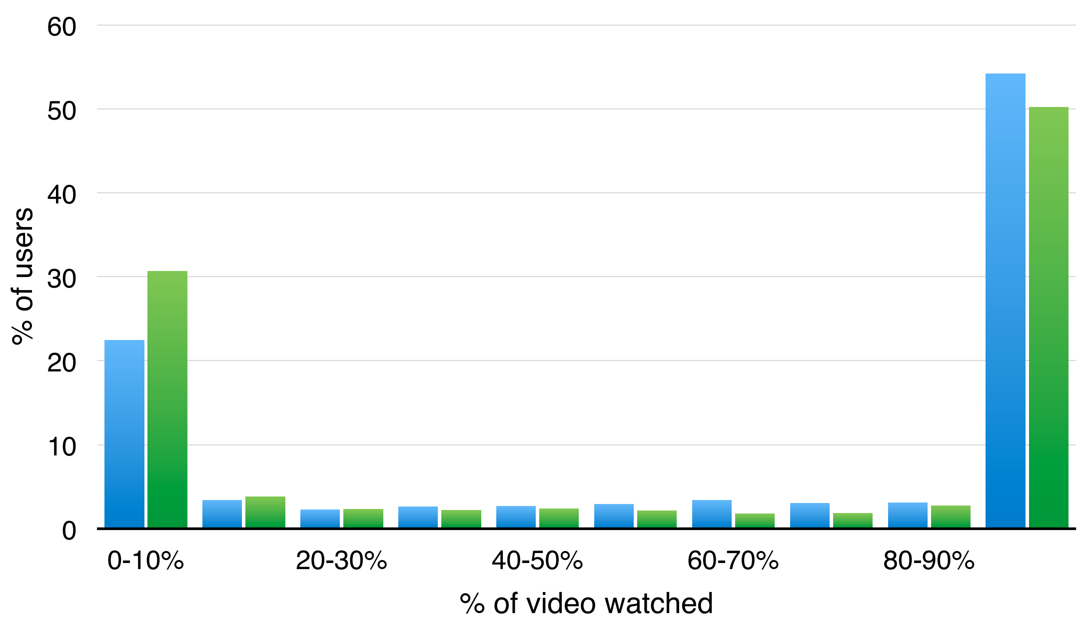

Figure 5. Graph of the total fraction of CS200 videos watched for two sets of users: working professionals (blue bars, on the left of each pair) and regular students (green bars, on the right).

To quantify the relationship between the frequencies at the two ends, we calculated the ratio of the percentage of users who watched 90-100% of a video and the percentage of users who watched 0-10% of a video. In other words, the higher the ratio, the more users watched 90-100% of the video relative to how many watched 0-10% of the video. Calculating this ratio for each video and taking the average across all videos, we found the average to be 3.92 for the working professionals and 2.26 for the regular university students, indicating some difference between the two groups.

In addition, we applied the same watching session analysis we used previously on the whole data set. Of the 17,034 total watching sessions for CS200, 14,973 were from regular students and 2,061 were from working professionals. The analysis found the median video watch time to be 13.9 minutes for the working professionals and 13.0 minutes for the regular students. And the median total time spent on a video during a watching session (including pauses) was 20.0 minutes for the working professionals and 19.5 minutes for the regular students.

Although there are clearly some differences between how the working professionals and regular students interacted with the videos, the data do not show a prominent difference between their video viewership distribution and watching behavior. This lack of significant difference may simply indicate that a good portion of the regular students skipped lectures and treated the course as a distance course. Another explanation could be that many of the regular students found it useful or necessary to view large portions of the videos in addition to attending lectures.

Conclusions

Tables 3 and 4 below summarize the main results of our analysis. The median minutes watched by a student during a single watching session was 12-13 minutes, with a mean in the range of 17-20 minutes (Table 3). But most students spent multiple watching sessions with each video. When all watching sessions by a student of an individual video were taken into account (Table 4), we found that a majority of students watched over 90% of each video in CS100 (a median of 45.7 minutes for each video) and over 70% of each video in CS200 (a median of 51.8 minutes for each video). (Results from the "Median Total % Viewed" and "Median Mins Viewed" columns in Table 4.)

	Mean Video Length (mins)	Median Mins Viewed	Mean Mins Viewed
CS100	50.1	12.1	17.4
CS200	66.4	13.0	19.5

Table 3. Amount of video watched by students in a single watching session.

	Mean Video Length (mins)	Median Mins Viewed	Mean Mins Viewed	Median Total % Viewed	Mean Total % Viewed
CS100	50.1	45.7	36.7	91.3	73.3
CS200	66.4	51.8	41.0	74.0	59.9

Table 4. Total amount of a video watched by students over multiple watching sessions. (Note that the Mean Total % Viewed will be equal to the Mean Mins Viewed divided by the Mean Video Length only when the lengths of all the videos are equal. This condition was true, or nearly so, for CS100. But for CS200 there was more variation in the video lengths. See Table 2 for the range of lengths.)

At a minimum we can therefore conclude that the "six-minute rule," while perhaps valid for a MOOC environment, does not capture the video viewing behaviors of students in more standard college courses that use videos. This does not necessarily mean that longer videos are equally as effective as shorter videos. There are still good reasons for instructors to think about how course material can be presented in shorter modules than the canonical 50-minute or 75-minute lecture. Based on our data from single watching sessions, a rule of thumb for maximum video length would be in the range of 12-20 minutes. But in general how students interact with online videos is more complicated than can be represented by a "however-many-minute rule."

Many other questions remain, of course. Chief among them is that in both MOOCs and blended courses we are guessing at the intention with which students load and watch videos. We can try to infer a user's intention from shorter skips or longer leaps, from short clicks or long pauses, but, really, we do not know. Going forward we plan to weave features into our user interfaces that help us collect data not just on how users are interacting with videos and other resources, but also why they are using them in that moment. In other words, rather than inferring intentions and contexts from interaction traces, we want to deduce them more directly through self-reporting or other means. We believe that this shift in focus for building online learning user interfaces can yield valuable insights as much about interface design as about the learning process.

A related challenge is how to disambiguate the various populations of students viewing online videos for a course. Why and how do different types of students interact with, or fail to interact with, online videos?[6] Is it possible to customize what a student sees based on prior users' viewing patterns? For example, if thousands of users watch a video, can the interaction peaks be used to create the equivalent of a sports or news highlight reel that focuses on key and/or challenging points?[12, 14, 15] Learning environments that include both online lectures and activities and in-person lectures and activities present very attractive targets for specifically these kinds of investigations. Though it depends on the design of the course, such learning environments often have students who never attend lectures and rely on videos to absorb material (i.e., similar to MOOCs) as well as students who always attend lecture and rely on videos for reference or review of course material (i.e., similar to offline materials). And perhaps the most important question of all, are there learning advantages or disadvantages associated with specific types of viewing behavior? We intend to address these questions as our studies continue.

Acknowledgements
We especially want to acknowledge the assistance of Ray Saray and Kyle Barnes of the Stanford Center for Professional Development. Their tireless work in extending the data gathering capabilities of the SCPD video player was crucial to the success of this project.

Bibliography

1. Breslow, L., D.E. Pritchard, J. DeBoer, G.S. Stump, A.D. Ho, and D. T. Seaton. 2013. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, *8*(1), 13-25.

2. Guo, P. J., J. Kim, and R. Rubin. 2014. How video production affects student engagement: An empirical study of MOOC videos. Paper presented at *L@S 2014*, March 4–5, 2014, Atlanta, Georgia, USA.

3. Kim, J., P.J. Guo, D.T. Seaton, P. Mitros, K.Z. Gajos, and R.C. Miller. 2014. Understanding in-video dropouts and interaction peaks in online lecture videos. Paper presented at *L@S 2014*, March 4–5, 2014, Atlanta, Georgia, USA.

4. Kim, J., P.J. Guo, C.J. Cai, S.W.D. Li, K.Z. Gajos, and R.C. Miller. 2014. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, pp. 563-572.

5. Kim, J., E.L. Glassman, A. Monroy-Hernández, and M.R. Morris. 2015. RIMES: Embedding interactive multimedia exercises in lecture videos. Paper submitted to CHI 2015 Conference, April 18-23, 2015 (acm.org), Seoul, Korea.

6. Kizilcec, R.F., C. Piech, and E. Schneider. 2013. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 170-179.

7. Li, F.C., A. Gupta, E. Sanocki, L.W. He, and Y. Rui. 2000. Browsing digital video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 169-176.

8. Matejka, J., T. Grossman, and G. Fitzmaurice. 2012. Swift: Reducing the effects of latency in online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 637-646.

9. Matejka, J., T. Grossman, and G. Fitzmaurice. 2013. Swifter: Improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1159-1168.

10. Medina, J. 2014. *Brain Rules*, pp. 74, 89-93.

11. Nguyen, C., Y. Niu, and F. Liu. 2012. Video summagator: An interface for video summarization and navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* pp. 647-650.

12. Olsen, D. R., and B. Moon. 2011. Video summarization based on user interaction. In *Proceedings of the 9th International Interactive Conference on Interactive Television*, pp. 115-122.

13. Ruedlinger, B. 2014. Does length matter? It does for video. http://wistia.com/blog/does-length-matter-it-does-for-video-2k12-edition/ (accessed 3 December 2014).

14. Shaw, R., and M. Davis. 2005. Toward emergent representations for video. In *Proceedings of the 13th annual ACM International Conference on Multimedia*, pp. 431-434.

15. Truong, B. T., and S. Venkatesh. 2007. Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *3*(1), 3.