Online Lecture Watch Patterns: A 3-Axis Model of Characterizing Students

Gina Hong
Student
University of British Columbia
ginahong@student.ubc.ca

Supervisor: Alison Myers
Research Analyst
UBC Sauder School of Business

Abstract

The global pandemic of COVID-19 has led to an unprecedented global lockdown that has forced most of learning to go entirely online. This has led to asynchronous online videos becoming our main vessel of delivery for lecture content. Given that the current state of research focuses mainly on MOOCs or hybrid courses or does not really attempt to characterize individual student video watch patterns, this paper attempts fill this gap. The paper proposes a 3-axis model of characterizing student video watch pattern, with the three axes/metrics being: playback speed, skipping frequency, and viewing consistency. The analysis was performed on two courses under UBC Sauder, and the data was collected via Panapto. Aggregation and calculations were performed to create a singular metric that each represents playback speed and skipping frequency, while several values were computed in order to better visualize viewing consistency. This research hopes to serve as a starting point for future research involving characterizing student watch data.

Keywords

Learning Analytics, Asynchronous Online Lecture, Video Watch Patterns, Instructional Design, Student Behavior, Statistical Analysis

1. Introduction

With the unprecedented global lockdown brought upon by COVID-19, the majority of schooling—at least in North America—is now online. As schools switch to online learning, asynchronous lecture videos have now been chosen as the main way to deliver course content. Combined with the granular data collection system built into many video hosting sites, there is an abundance of student video watching data that can now be analyzed.

However, perhaps due to the relative recency of both COVID-19 and our following reliance on asynchronous lecture videos, there is surprisingly little research that attempts to calculate a 'metric' or a 'model' in which we can characterize student lecture watching habits.

This paper proposes and explores a potential 3-axis model that can be used to better understand student video watching data. It characterizes student video habits via three axes: playback speed, skipping frequency, and viewing consistency. This 3-axis model was explored by analyzing anonymized student video watching data, collected from two courses under UBC Sauder, where the lecture videos were hosted via Panapto.

One foresees two particular advantages of this 3-axis model of student video watching habits. First, by quantifying certain aspects of online video watching—like skipping, playback speed—it can more easily communicate to content developers the effectiveness of their lecture videos. Especially when considering how granular the raw data tends to be, it is essential that there is some framework that can represent student interaction in a more approachable and digestible manner. Second, by clarifying how students are interacting with videos, it provides a guideline for educators when they create future videos; this can hopefully increase student engagement with asynchronous online lectures.

2. Literature Review

The field of learning analytics has several branches and active research regarding online education. The particular subsection of research that relates to this paper focuses on video watching habits. A comprehensive literature review showed that the research on this topic can be divided under two categories: Massive Open Online Courses (MOOCs) or Hybrid courses.

2.1 Video watching habits: MOOCs

Li et al. (2015) linked student video interactions with perceived difficulty of the course content and student course performance for a particular MOOC. The study found that the easier a student perceived the course to be, the more likely the student was to skip through the videos. Additionally, students who performed better in the course were less likely to interact (i.e., skip through, rewind) with the course video.

Guo et al. (2014) and Kim et al.'s (2014) research utilized the same data from undergraduate MOOCs to discuss student engagement time with video completion rates. The researchers revealed that median engagement time for students was 6 minutes, and students tended to not complete longer videos. In particular, Kim et al.'s research identified and ranked frequent "actions" as: starting a new lecture video, continuing from a previous session, completing tutorial videos, and so on.

Although research in video watching habits with MOOCs shed a significant amount of info on how students utilize online videos, it would be faulty to assume that these tendencies directly match up with university courses that utilize asynchronous online videos. MOOCs tend to vary largely in their student demographic and experience of students, while also having a significantly lower course completion rate of 10% (Ahn and Bir, 2018). As such additional literature review involving university hybrid courses was conducted.

2.2 Video watching habits: Hybrid Courses

Similar to this paper, research on student video watching habits in hybrid courses utilize data from video management systems. A common result throughout multiple researches seemed to be a “peaking” of student views immediately before exams (Kinsella, Mahon, and Lillis, 2017; Elliot and Neal, 2016).

Some research also focused on video completion and usages in hybrid courses. Using data from an economics course, Elliott and Neal (2016) found that while views peaked right before exams, students tended to only watch portions of lecture videos. This tendency to only watch portion of online videos was not limited to this study. A study by Gross and Dinehart (2016) revealed that only half of the students watched at least 53% of the pre-lecture videos, and even these numbers declined as the course progressed.

These results show valuable student watch pattern and tendencies with online videos. However, while the hybrid university courses and online asynchronous courses are alike in demographic, student experience, and course dropout rates, hybrid courses still differ from asynchronous courses in that significant portion of the learning is conducted in person, surrounded by instructors and peers. Additionally, the research mentioned so far do not attempt to characterize or categorize individual student’s online lecture viewing habits. Thus, this research hopes to fill in this gap in both asynchronous online lectures and in characterizing individual student watch patterns.

3. Method

Table 1. *Anonymized raw data containing course video information*

	FolderId	FolderName	SessionId	SessionName	Duration
0	55a9M66B-2aBM-2B1x-aB5B-ax2aMB6xf8B9	Course_A	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	Video 0	182.336
1	55a9M66B-2aBM-2B1x-aB5B-ax2aMB6xf8B9	Course_A	6BB61f82-9522-2a82-9ab8-ax2xMMMb82xB	Video 1	752.981
2	55a9M66B-2aBM-2B1x-aB5B-ax2aMB6xf8B9	Course_A	9M522189-26x9-2219-9x52-axB2MB222869	Video 2	2886.842
3	55a9M66B-2aBM-2B1x-aB5B-ax2aMB6xf8B9	Course_A	2xb89a52-1Bxa-2xxx-86Bx-ax28MB6916fx	Video 3	217.221
4	55a9M66B-2aBM-2B1x-aB5B-ax2aMB6xf8B9	Course_A	a5x6a265-15Bf-2Bf6-bBBf-ax29MB6bxM8M	Video 4	393.785

Table 2. *Anonymized raw data containing student watch data*

	SessionId	UserId	Date	DateTime	PlaybackSpeed
0	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	2020-11-17	2020-11-17 16:55:25.093000+00:00	1.0
1	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	2020-11-17	2020-11-17 16:55:25.100000+00:00	2.0
2	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	2020-11-17	2020-11-17 16:56:54.977000+00:00	2.0
3	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	2020-11-17	2020-11-17 16:57:25.040000+00:00	2.0
4	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	2020-11-17	2020-11-17 16:57:53.983000+00:00	2.0

	SessionId	UserId		StartPosition	StartReason	StopPosition	StopReason	SecondsViewed
0	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB		0.000000	Start	12.423785	Pause	12.423785
1	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB		12.513214	NewSpeed	132.953073	Pause	120.439859
2	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB	...	133.313303	Resume	153.346660	Pause	20.033357
3	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB		153.554968	Resume	171.014762	Pause	17.459794
4	8829Mx52-x5bB-2219-895M-ax2xMMMb2xaB	912bBM22-x91M-2x5M-96f8-ax2xMB8BxBxB		171.362516	Resume	173.852785	PlayerClose	2.490269

3.1 Raw data

In order to address the 'how-to' and the reasoning behind each individual metrics, one must discuss the initial raw data. The data was collected from two courses— anonymized as course A and course B—where the lecture videos were hosted via Panapto. The data was delivered in two csv files: course video information and student activity data.

The course video information data was quite simple and contained the 'folderId' (equivalent to courseID), 'SessionId' (equivalent to videoID), and duration of the video. Table 1 shows the first few rows of this video information data for course A.

The main data for this analysis came from the student watch data. Table 2 shows the first few rows of student watch data for course A. In this data, each row represents a 'click event/interaction' involving a particular video by a student. Each row records the videoID, studentID, exact date, playback speed, and start/stop reason for the click event. This incredible granularity of the data meant that various aggregation methods and exploratory analysis had to be completed in order to understand the data to determine the research direction. Jupyter Notebook, Pandas Python library, and Tableau was used during the data exploration phase. Following the data exploration phase, the three axes to be explored were determined to be: playback speed, skipping frequency, and viewing consistency.

3.2 Calculating playback speed metrics

The playback speed aims to represent the overall speed in which students viewed the asynchronous lecture videos. This was measured as a value called 'avg_pb_speed' and is a weighted average of the 'PlaybackSpeed' field seen in Table 2. By accounting for the duration (i.e., a particular click event's SecondsViewed) in proportion to the total secondsViewed for a particular video, the metric avoids short or insignificant outlier playback speeds from skewing the data.

$$avg\ pb\ speed = playbackSpeed\ * \frac{eventDuration}{SUM(eventDuration)}$$

The initial field was calculated by grouping the raw data by user and video. After identifying the average playback speed for each user and video, an overall course average of this value for each user was calculated and selected as the user's overall playback speed for the course.

3.3 Calculating skipping metrics

The skipping frequency refers to the student's tendency to jump from one part of a video to another instead of continuously watching. This was calculated by dividing the number of skip events with the total number of click events for a particular user on a particular video. This means for each video, a particular user's skipping percentage was first calculated. A click event or row was denoted as a skip event if either one of the StartReason or StopReason value was "Seek".

$$skipPercentage = \frac{COUNT(skip\ events)}{Total\ number\ of\ events}$$

Once the skipping frequency was calculated for each user and video, the user's overall skipping frequency over the whole course was calculated via a group by operation.

3.4 Aggregating for viewing consistency

Unlike the metrics for playback speed or skipping frequency, viewing consistency proved to be more difficult of a trait to represent as one value. Not only did one have to consider the dates of student activity, but one also had to consider the length of the activity (i.e., how long the student viewed the lecture video). The idea of viewing consistency involves how consistent/regular a student engaged with the course and the overall 'intensity' of the engagement.

Table 3. *Sample rows from viewing consistency table*

	Date	UserId	SessionId	...	totalWatchTime	vidProgress	totalWatchPercent
0	2020-10-18	11fx52B6-x1Bb-26Bx-98xx-ax22MB221x2x	8x2MbxB1-M92b-2x9B-85xa-ax5xMB62B8x5		16080.091744	0.713197	0.122439
1	2020-10-18	11fx52B6-x1Bb-26Bx-98xx-ax22MB221x2x	B9bbx2a6-baxM-26fB-852B-ax58MB2f9519		16080.091744	0.015410	0.003127
2	2020-10-19	11fx52B6-x1Bb-26Bx-98xx-ax22MB221x2x	8x2MbxB1-M92b-2x9B-85xa-ax5xMB62B8x5	...	16080.091744	0.124227	0.021327
3	2020-10-19	11fx52B6-x1Bb-26Bx-98xx-ax22MB221x2x	96b2189b-xx2a-2xbB-a19f-ax58MB6x2Bx9		16080.091744	0.065047	0.008323
4	2020-10-19	612b5xBM-BaM1-2xx1-a8a2-axMfMBxxxaB9	8x2MbxB1-M92b-2x9B-85xa-ax5xMB62B8x5		92.865769	0.003439	0.102222

As such, each row in the viewing consistency table represents a student's progress on a particular video on a particular day. The progress is represented via two columns, 'vidProgress' and 'totalWatchPercent'. 'totalWatchPercent', in particular, shows the percentage of video watch time in relation to the overall watch time spent by the student for the entire course. Although viewing consistency was not represented as a single value, my hope was that these set of records could be used to visualize viewing consistency and setup for future research in this particular area.

4. Results

4.1 General overview of data

Table 4. *Descriptive statistics of average video length, number of videos, student count, min/max dates for each Course*

	FolderName	avg_vid_length	total_videos	total_runtime	student_cnt	avgSec_per_row	min_date	max_date
0	Course_A	898.033104	96	86211.178	3013	47.990185	2020-09-02	2021-02-02
1	Course_B	2467.309567	30	74019.287	2537	63.323545	2020-10-18	2021-02-03

The analysis was conducted on student watch data from two courses: course A and B. Table 4 shows the overview statistics for each course. Course A had significantly more videos at a total video count of 96, in comparison to course B's 30. But course B had lengthier videos averaging

~40 minutes; this difference is a point of interest, which can hopefully be discussed as we compare the metrics between these two courses. Both courses ran sometime during the 2020W term 1 and had a fairly similar student count of 2500+ unique userIDs.

4.2 Playback speed

Table 5. *Descriptive statistics of playback speed metric for course A and B*

	FolderName2	count	mean	std	min	25%	50%	75%	max
0	A	3013.0	1.227733	0.282660	0.0	1.0	1.097231	1.400396	2.0
1	B	2537.0	1.212058	0.282904	0.0	1.0	1.020576	1.409861	2.0

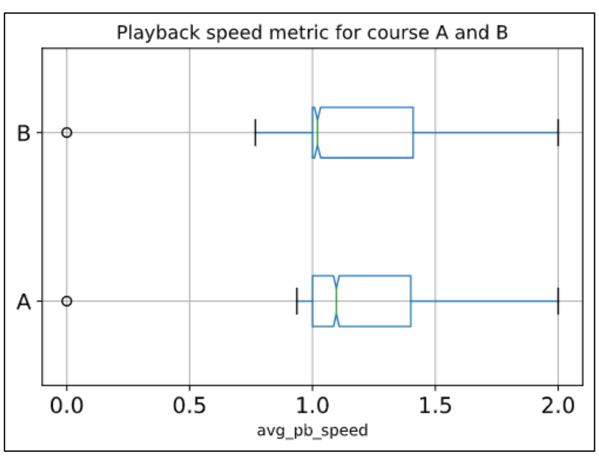
Table 6. *Descriptive statistics of total watch time for course A and B*

	FolderName2	count	mean	std	min	25%	50%	75%	max
0	A	3013.0	12457.331525	18386.178921	1.003691	559.321803	2415.058578	20329.313317	144918.005186
1	B	2537.0	5234.388647	10729.490183	1.012118	205.082830	1286.082918	3028.412806	90684.169614

In order to calculate the metric for playback speed (avg_pb_speed), the total watch time for each student was calculated. The descriptive statistics for both columns— 'avg_pb_speed' and 'totalWatchTime'—is represented via Table 5 and 6. The mean playback speed was consistent for both courses at approximately 1.2x, perhaps suggesting that there is some playback speed that becomes the average value regardless of the course.

Some interesting numbers to note are the numbers related to total watch time. In table 4, one could see that the total runtime of course A was longer than course B. This difference seems to have impacted the total watch time of students, leading to significantly higher watch time values for course A in both mean, median, and 75% values. The fact that the median total watch time value is not even half of the total runtime of the courses may be due to the current assumption that each unique userID represents a unique student. Some students may be utilizing several devices to interact with the course, which would lead to lower total watch time means/median/etc.

Figure 1 [Left]. *Boxplot of playback speed metric for course A and B*



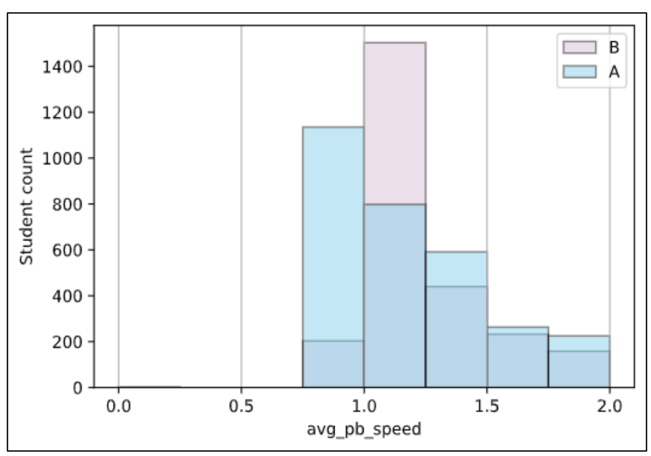


Figure 2 [Right]. *Histogram of playback speed metric for course A and B*

The histogram drawn for average playback value reveals that most students in course A chose a playback speed between [0.75, 1], while most students in course B chose a playback speed between [1, 1.25]. There seems to be a link between the average video length and playback speed. Course B had a significantly longer average video length at 2467 seconds, compared to course A's 898 seconds. Perhaps students were more likely to choose a faster playback speed when confronted with a lengthier video.

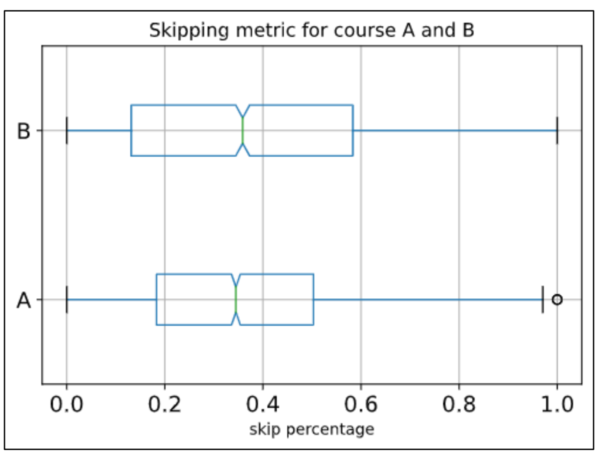
4.3 Skipping

Table 6. *Descriptive statistics of skipping percentage for course A and B*

	CourseName	mean	min	25%	50%	75%	max
0	A	0.351964	0.0	0.183292	0.344828	0.503198	1.0
1	B	0.374510	0.0	0.131579	0.358696	0.583333	1.0

The descriptive statistics for skipping revealed a slightly higher skipping percentage for students in course B compared to course A. However, considering that the 25% value for course B is significantly lower than course A and the wider box for course B's boxplot, it seems that course B saw a wider range of skipping habits shown by its students. Additional to the wider range of skipping habits, it seems that while the greatest number of students were placed within a skipping percentage smaller than 0.2(20%) for course B, the greatest number of students for course A were within a skipping percentage of [0.2, 0.4]. These differences in skipping tendencies between the two courses could be explained by the difference in average video length between course A and B or by the information density in the videos.

Figure 3 [Left]. *Boxplot of skipping percentage for course A and B*



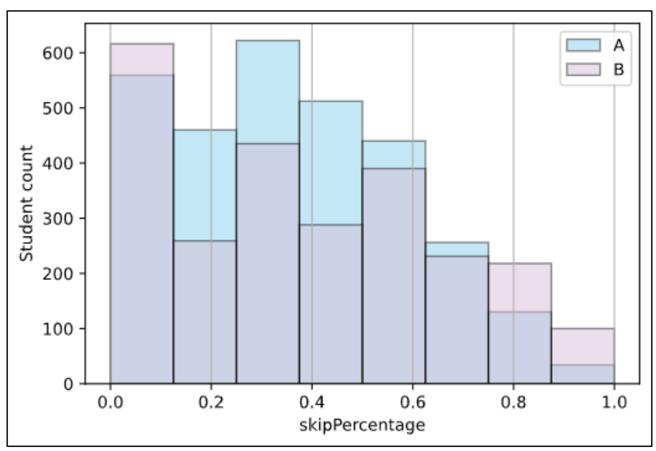


Figure 4 [Right]. *Histogram of skipping percentage for course A and B*

4.4 Viewing Consistency

Figure 5 [Left]. *Scatterplot of each student's daily viewing percentage for course A*

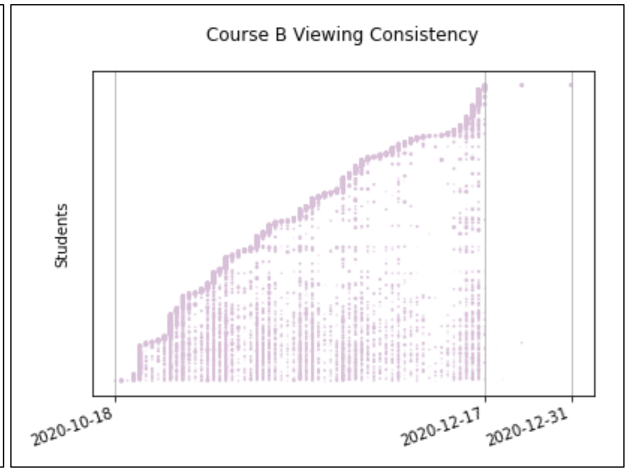


Figure 6 [Right]. *Scatterplot of each student's daily viewing percentage for course B*

As mentioned in section 3.4, calculating a single metric for viewing consistency proved to be difficult as this metric would have to summarize and represent data from two different angles: the regularity in which student engaged with the course (i.e., dates) and the overall length of each engagement (i.e., watch time). To get around this difficulty of creating a single metric, an attempt to visualize the viewing consistency led to a scatter plot seen in Figure 5 and 6 above.

Each 'line'/y-axis of the scatterplot represents a unique user/student. The size of each dot on the scatterplot reflects the totalWatchProgress metric (i.e., percentage of video watched that day in relation to overall time spent on course). These dots are plotted by the corresponding dates. Thus, the scatterplot visualizes the overall engagement pattern by the students.

5. Discussions

5.1 Limitations of viewing consistency

The scatterplots shown in Figure 5 and 6 visualizes the overall engagement pattern by each student. A denser line that runs from the beginning of the course to the end of a course, would indicate a consistent student. On the other hand, a short but thicker line would indicate a 'crammer' student watch pattern. Additionally, lengthy but 'thin' lines could indicate a student who didn't fully watch the videos although they engaged with the course consistently.

For the scope and timeline of this research, I was not capable of devising a formula that could represent this viewing consistency line/pattern. If future research is done to create a metric that represents both the density, thickness, and length of this viewing consistency line, that metric could be used to represent viewing consistency.

Despite the absence of a singular metric that represents each student's viewing consistency, the scatterplot does reveal a certain 'synchrony' to when students tended to watch videos. The fairly notable vertical lines probably indicate that students tended to watch particular lecture videos at the same time. The dimness or sparseness of course B's scatterplot also reveals some interesting traits about course B that is quite different from course A. The overall sparseness in course B's scatterplot as the date approaches 2020-12-17, probably indicates that many students fell behind on watching videos; an interesting pattern that perhaps relates to the lengthier average video length of course B.

5.2 Potential for exploring metric combinations

With the metrics related to playback speed and skipping frequency calculated and aggregated for each student, this opens the door for exploring the combinations/relationship between playback speed and skipping frequency. This could help in answering questions such as: "are students who watch videos at a higher speed more likely to skip through videos? (i.e., are 'impatient' students a potential archetype of online lecture video watching?)".

Additionally, if a viewing consistency metric can be devised based on the aggregated viewing consistency table completed by this research, a more complex combination-check can be performed. By characterizing a student's watch pattern on 3 axes, one would be able to identify relationships between playback speed, skipping frequency, and viewing consistency: hopefully leading to a richer characterization of each student.

6. Conclusion

As COVID-19 drove the majority of education online, our reliance on asynchronous online lecture videos to deliver course content has grown exponentially. This, in combination with the granular data recording provided by content hosting sites, has opened many potential doors for exploring student lecture video watch patterns/habits. However, due to the recency of COVID-19, literature review showed that the current state of research focused 1) on either MOOCs or hybrid courses or 2) did not attempt to characterize or categorize individual students. As such, this research embarked as an attempt to propose a set of metrics that could characterize student video viewing habits for asynchronous online courses.

Through aggregating on courseID, userID, and date, singular metrics were calculated for playback speed and skipping tendencies. This revealed interesting differences and similarities between courses A and B that could be explained by differences in average video length or information density of the videos. Additionally, values were computed in order to better visualize student viewing consistency. While the scope of this research was not enough to finalize a singular metric to represent viewing consistency nor perform a cluster analysis using the three metrics, it is my hope that the results so far has 1) successfully illustrated the usefulness of having a set of metrics to characterize student video watch patterns and 2) can provide a starting point for further research in this topic.

References

Ahn, Benjamin, and Devayan D. Bir. “Student Interactions with Online Videos in a Large Hybrid Mechanics of Materials Course.” *Advances in Engineering Education*, vol. 6, no. 3, 0 2018, https://eric.ed.gov/?id=EJ1181575.

Dazo, Suzanne L., et al. “An Empirical Analysis of Video Viewing Behaviors in Flipped CS1 Courses.” *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*, Association for Computing Machinery, 2016, pp. 106–111, doi:10.1145/2899415.2899468.

Elliott, Caroline, and David Neal. “Evaluating the Use of Lecture Capture Using a Revealed Preference Approach.” *Active Learning in Higher Education*, vol. 17, no. 2, July 2016, pp. 153–67, doi:10.1177/1469787416637463.

Gross, Shawn P., and David W. Dinehart. 2016. “Pre- and Post-class Student Viewing Behaviors for Recorded Videos in an Inverted Sophomore Mechanics Course.” Paper presented at the ASEE Annual Conference & Exposition, New Orleans, Louisiana.

Guo, Philip J., et al. “How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos.” *Proceedings of the First ACM Conference on Learning @ Scale Conference*, Association for Computing Machinery, 2014, pp. 41–50, doi:10.1145/2556325.2566239.

Kim, Juho, et al. “Understanding In-Video Dropouts and Interaction Peaks Inonline Lecture Videos.” *Proceedings of the First ACM Conference on Learning @ Scale Conference*, Association for Computing Machinery, 2014, pp. 31–40, doi:10.1145/2556325.2566237.

Kinsella, Gemma K, Catherine Mahon, and Seamus Lillis. 2017. “Using Pre-Lecture Activities to Enhance Learner Engagement in a Large Group Setting.” Active Learning in Higher Education 18, no. 3 (June): 231–42. https://doi. org/10.1177/1469787417715205.

Li, Nan, et al. “MOOC Video Interaction Patterns: What Do They Tell Us?” *Design for Teaching and Learning in a Networked World*, edited by Gráinne Conole et al., Springer International Publishing, 2015, pp. 197–210, doi:10.1007/978-3-319-24258-3_15.