Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos

Juho Kim[1] Philip J. Guo[2] Daniel T. Seaton[3] Piotr Mitros[4] Krzysztof Z. Gajos[5] Robert C. Miller[1]

[1]MIT CSAIL
{juhokim, rcm}@mit.edu

[2]University of Rochester
pg@cs.rochester.edu

[3]Office of Digital Learning, MIT
dseaton@mit.edu

[4]edX
pmitros@edx.org

[5]Harvard SEAS
kgajos@eecs.harvard.edu

ABSTRACT

With thousands of learners watching the same online lecture videos, analyzing video watching patterns provides a unique opportunity to understand how students learn with videos. This paper reports a large-scale analysis of in-video dropout and peaks in viewership and student activity, using second-by-second user interaction data from 862 videos in four Massive Open Online Courses (MOOCs) on edX. We find higher dropout rates in longer videos, re-watching sessions (vs first-time), and tutorials (vs lectures). Peaks in re-watching sessions and play events indicate points of interest and confusion. Results show that tutorials (vs lectures) and re-watching sessions (vs first-time) lead to more frequent and sharper peaks. In attempting to reason why peaks occur by sampling 80 videos, we observe that 61% of the peaks accompany visual transitions in the video, e.g., a slide view to a classroom view. Based on this observation, we identify five student activity patterns that can explain peaks: starting from the beginning of a new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation. Our analysis has design implications for video authoring, editing, and interface design, providing a richer understanding of video learning on MOOCs.

Author Keywords

Video analysis; in-video dropout; interaction peaks; online education; MOOC; peak detection.

ACM Classification Keywords

H.5.1. Information Interfaces and Presentation (e.g. HCI): Multimedia Information Systems: Video

INTRODUCTION

MOOCs often include hundreds of pre-recorded video clips. Recent research on the first edX course, 6.002x, has shown that learners spend a majority of their time watching videos [2, 23], but little research has been aimed at the click-level interactions within MOOC videos. With thousands of learners watching the same online lecture videos, video analytics can provide a unique opportunity in understanding how learners use video content and what affects their learning experience.

This paper analyzes click-level interactions resulting from student activities within individual MOOC videos, namely playing, pausing, replaying, and quitting. We analyze video player interaction logs from four MOOCs offered on the edX platform to identify temporal interaction patterns at the second-by-second level. Specific focus is given to 1) in-video dropout rates and 2) peaks associated with re-watching sessions and play events.

Video dropout, i.e., navigating away from a video before completion, is a measure of engagement. With more engaging videos students might stay until later in the video, resulting in lower dropout. Instructors using videos in their pedagogy need to know what aspects of their videos are the most engaging or most widely viewed. While existing analytics tools provide access to this data, they do not consider different video kinds (lecture or tutorial) and presentation styles (slides, head shot, etc.) specific to the educational context.

When a significant number of students interact with a common portion of a video, the resultant data can be binned to highlight *peaks* in the video timeline. **Peaks in viewership and student activity** can precisely indicate points of interest for instructors and students. These spikes, hereinafter referred to as interaction peaks, can indicate student confusion, introduction of important concepts, engaging demonstrations, or video production glitches. We manually inspected 80 videos from our set to understand why these peaks occur. One notable observation we made is that the peaks often coincide with visual transitions in a video, such as switching from a slide to a classroom view, or from handwritten notes to a software screencast. Combining the interaction data with visual content analysis, we identified five student activity types that can lead to a peak.

This paper makes the following contributions:

- A first MOOC-scale in-video dropout rate analysis, finding higher dropout rates in longer videos, re-watching students (vs first-time watchers), and tutorials (vs lectures).

L@S 2014, March 04–05 2014, Atlanta, GA, USA.

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

- A first MOOC-scale in-video interaction peak analysis, finding more frequent and sharper peaks in re-watching students (vs first-time watchers) and tutorials (vs lectures).
- Categorization of student activities responsible for a peak: starting from the beginning of a new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation.
- Data-driven design implications for video authoring, editing, and interface design in the context of MOOCs that reflect the temporal interaction patterns of students.

In the remainder of the paper, we discuss related work and our analytical measures and methods. We then report results from the in-video dropout and interaction peak analysis, and introduce five activity categories that might be a cause of a peak. We present design implications for better video learning experiences, and conclude with limitations and future work.

RELATED WORK

Existing research on video engagement analysis has involved three general methods: implicit user data (interaction log), explicit user data (clicking the "important" button, voting), and content analysis (visual, speech, or transcript analysis).

First, implicit user data has the benefit of requiring no additional action on user's part, because this data is automatically captured by the system while users naturally interact with videos. Shaw and Davis advocate using actual viewership data in modeling user interest [24]. Existing systems leverage scrubbing [29], zooming and panning [3], and playing and pausing [4] activities. SocialSkip [4] demonstrates that modeling users' video interactions can accurately capture user interest in information retrieval tasks. While our work adopts the idea of using video clickstream data from the literature, our analysis differs in that it uses large-scale interaction data from MOOCs, and that it focuses on in-video dropout and interaction peaks in the educational context.

Secondly, explicit user data can be collected by asking users to make a specific action around their points of interest. Previous research used user rating data [20] or annotations [24]. CLAS [21] is a lecture video annotation tool where students click a button when they find a part of the video important. The system aggregates responses from all students in a class to visualize important points. Deploying CLAS-like systems at MOOC-scale will provide useful complementary data to implicit user logs.

Content-based video analysis [25] has long been an active research area. Previous research uses image analysis and computer vision to extract keyframes [8], shot boundaries [18], or visual saliency [11]. We add a simple pixel difference metric to our analysis, and plan to incorporate more advanced techniques in future work. In summary, to the best of our knowledge, this work is a first MOOC-scale analysis for videos that combines interaction data and content-based analysis.

Tools for temporal pattern analysis

Understanding temporal patterns in large-scale video data requires powerful computational and visual tools. We present existing research and systems for each.

Temporal pattern analysis of time-series data inspired the analytical methods used in this work. Kleinberg [16] introduced a burst model for detecting meaningful structure in documents, and Jones and Diaz [12] applied this model among other temporal features to identify temporal patterns in search queries. Using search query and social media streams, researchers categorized search query patterns and trending events based on the shape of spikes [17, 13]. This paper applies similar techniques to analyze video interaction patterns, which is enabled by large-scale student data collected from MOOCs.

Video analytics platforms can enable the visual sensemaking of large-scale data. General purpose video platforms such as Youtube provide advanced analytics [9, 28] for content authors. These services include dashboards showing viewership graphs over time for a video, and suggest focusing on rises and dips. Our analysis considers more in-depth activity data such as play, pause, and skip events on the player, and content specific to educational videos, such as video kinds (lecture or tutorial), presentation styles (slide, head shot, etc.), and visual transitions between the presentation styles.

VIDEO INTERACTION DATASET

Our dataset consists of interaction logs from the edX video player over four courses offered in Fall 2012. Each log entry contains user name, time of access, video ID, event type, and internal video time, as documented in [7]. A play event is created when the user clicks the play button on the player or scrubs the playhead to a new position while the video is playing. A pause event is created when the user clicks the pause button or scrubs the playhead to a new position when the video is paused.

Table 1 summarizes information on the four courses and their videos. We chose the courses offered at roughly the same time to minimize the effect of changes in the edX platform, logging method, and student population. They span different institutions, subject fields (computer science, statistics, or chemistry), and recording styles. One of the authors manually labeled video types and presentation styles for all the videos in the video set. Video types represent a pedagogical purpose of a video, including introduction, tutorial, or lecture. Presentation styles represent the visual format of instruction: Powerpoint-style slide, code editor, head shot, classroom recording, and handwritten tablet annotations similar to those used in Khan Academy videos.

Data Processing Pipeline

Our data processing pipeline first reconstructs the watching history of each viewer and then aggregates the per-viewer history data to produce activity statistics for each second-long segment of the video. Specifically, the first step converts raw interaction log entries into watching segments. A watching segment keeps track of all continuous chunks of a clip watched by a user. It includes start and end time for every

Course	Subject	University	Students	Videos	Video Length	Processed Events
6.00x	Intro. CS & Programming	MIT	59,126	141	7:40	4,491,648
PH207x	Statistics for Public Health	Harvard	30,742	301	10:48	15,832,069
CS188.1x	Artificial Intelligence	Berkeley	22,690	149	4:45	14,174,203
3.091x	Solid State Chemistry	MIT	15,281	271	6:19	4,821,837
Total			**127,839**	**862**	**7:46**	**39,319,757**

Table 1. Overview of the four edX courses in our dataset offered in Fall 2012. "Students" refers to the number of students who watched at least one video, "Videos" is the number of all video clips posted, "Video Length" is the mean duration, and "Processed Events" is the number of total play and pause events captured by the video player.

watched segment. The second step uses the segment information to create second-by-second counts of viewers, unique viewers, re-watching sessions, play events, and pause events. Re-watching sessions only consider a student watching a segment of a video twice or more. Play and pause events increment a bin count if the event is triggered within that bin. Finally, such information can be queried upon request for statistical analysis and further processing.

The data processing module was implemented using Insights, the open source learning analytics library [6], which supports streaming events over SOA (Service-Oriented Architecture) as well as handling requests for query and view. It also uses Python, MongoDB, and the d3 visualization library [1].

ANALYSIS 1. IN-VIDEO DROPOUT

A dropout rate is defined by the percentage of students who start watching a video but leave before the video finished playing entirely. The dropout rate can reveal the factors that affect students to leave a video, helping video authors to consider them. Also, comparing this rate between videos can illustrate the relative difference in engagement. This analysis could provide valuable feedback to content creators whose courses are rapidly moving toward flipped environments where content consumption occurs online. To our knowledge, no previous work has studied the dropout rates within individual MOOC videos.

Method

For a video of length *n* seconds, let *viewcount(t)* denote the number of unique viewing sessions that include this second for each video. We compute the dropout rate of all videos in our set as: *1.0 - viewcount(n) / viewcount(0).* Note that all edX videos automatically start playing once the page is open, which might affect the results.

Results

On average across all videos, about 55.2% of viewing sessions (std=14.7) were dropouts before the end. Out of the 55.2% that dropped out, 36.6% (std=11.1) occurred within the first 3% of the video length. This means that 18.6% of the dropouts occur during the rest of the length. It is notable that the dropout rate changes quite dramatically at the beginning of a video.

Why do so many students leave the video very early on? The student might have left the video shortly after it (auto-)started, or the auto-play feature in the edX video player inadvertently started a video. Misleading video titles or course navigation interfaces might be another reason. A tip for content owners on YouTube analytics [9] states that viewers leaving before 5-10 seconds probably means the video keyword or title might not accurately represent the content. Additional analysis looking at the common page navigation paths of these early-dropping students might reveal issues with the video title or course navigation structure.

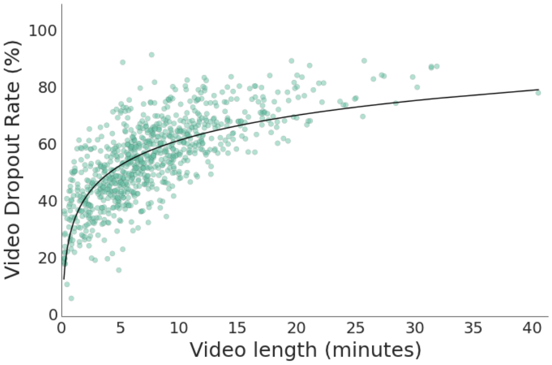


Figure 1. Longer videos exhibit higher dropout rates. Our linear regression model uses the log-transformed video length (x-axis) to predict the dropout rate (y-axis). The model fits the data well with r=0.55 with 95% CI = [0.50, 0.59].

The dropout rate increases with video length (Figure 1). Linear regression shows that the logarithmic value of the video length significantly predicted the dropout rate (b = 0.13, t(848) = 32.22, p <001). The overall model with the logarithmic value of the video length also predicted the dropout rate very well (adjusted R^2 = 0.55, F(1, 848) = 1038, p <0.001). This suggests that for a five-minute video, the predicted dropout is 53% (35% in the first 3%), whereas for a 20-minute video the rate goes up to 71% (47% in the first 3%). With longer videos, students might feel bored due to a short attention span or experience more interruption.

A recent analysis of edX data [10] shows that learner engagement drops significantly if the video length is longer than 6 minutes. Their analysis differs from ours in that they use viewing session length as engagement, as opposed to second-by-second dropout rates. Our analysis can provide additional evidence to the finding that shorter videos are more engaging because more students would drop out.

Another factor that might affect the dropout rate is whether the student watches the video for the first time. Students that are re-watching a video might have more specific information

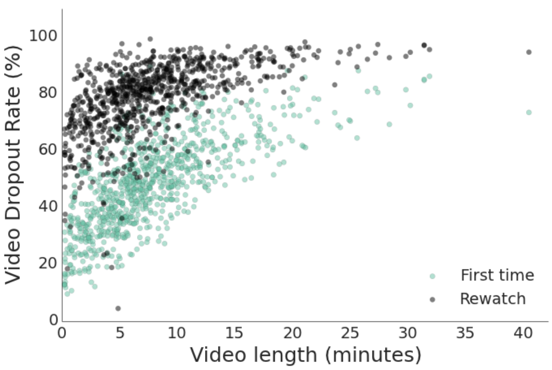


Figure 2. Re-watching students tend to drop out more, which might mean that re-watching students watch videos more selectively with a more specific need.

needs and selectively watch a video. Our analysis verifies this assumption as can be seen in Figure 2: the dropout rate of re-watchers (78.6%, std=54.0) was much higher than that of first-time watchers (48.6%, std=34.0). A Mann-Whitney's U test shows a significant effect (Z = -30.7, $p < 0.001$, r = 0.74).

Finally, we look at how video production types affect the dropout rate by comparing lecture videos and tutorial videos. Tutorial videos showed higher dropout rate (61.3%, std=38.3) than lecture videos (54.1%, std=36.3). A Mann-Whitney's U test shows a significant effect (Z = -5.29, $p < 0.001$, r = 0.18). One explanation is that lecture videos contain first-time introductions to concepts and sequential flow, whereas tutorial videos contain step-by-step instructions students can selectively review and follow along. The mean video length was not significantly different between the two video types ($p > 0.05$), limiting the effect of video length in the result.

ANALYSIS 2. INTERACTION PEAKS

In addition to staying in a video or leaving, students also actively play, pause, or skip the video to learn at their own pace. Uncovering meaningful patterns from these natural learning activities can provide an in-depth look at video learning on MOOCs. The temporal profiles of such patterns reveal time-specific interest, which might indicate student confusion, pacing issues in the video, useful information presented visually, or important concepts. Course instructors can refer to such information to attend to specific parts of a video. Comparing peak profiles between pedagogically different videos (lecture vs tutorial) can reveal the difference in students' consumption patterns, while comparison between watching contexts (first-time vs re-watching) might highlight different purposes in watching videos.

We investigate **temporal peaks** in the number of interaction events in particular, where a significantly large number of students show similar interaction patterns during a short time window. We use the following two peak definitions.

- A **re-watching session peak** is a sudden spike in the number of re-watching sessions during a period inside a video. We exclude first-time sessions because they tend to be more sequential. We instead focus on non-sequential, random access activities. Note that this measure is not per unique student. A student repeatedly watching a part of a video five times adds five to our measure.

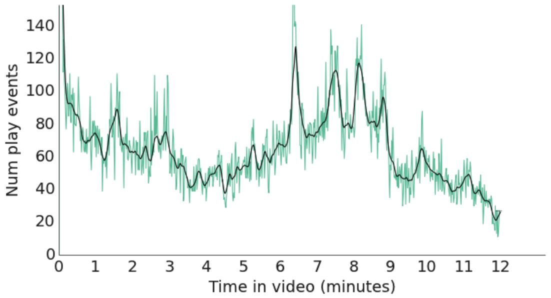


Figure 3. Even after aggregating data into bins of one second, the data is noisy (green curve). Kernel-based smoothing reduces noise in the data and helps salient patterns stand out (black curve).

- A **play event peak** is a sudden spike in the number of play events on the player. These events occur when a student clicks the play button or scrubs the playhead to a new position. We ignore autoplay events at the beginning of a video because they do not represent student-initiated activity.

Method

Raw watching session and interaction data are noisy (green curve in Figure 3). Identifying peaks in such noisy data both manually and automatically becomes difficult due to local maxima and false peaks. Following the bin-summarize-smooth framework [27], we first bin the data into one-second segments, which simplifies the computation and visualization. We then count all points in each bin to represent an aggregate number of events in a bin. To fight the noise and excessive variance in data and compensate for lost statistical power, we then apply smoothing to the binned and aggregated data (black curve in Figure 3). The smoothing technique we use is lowess (locally weighted scatterplot smoothing) [5], with the smoothing parameter of 0.02 after testing various values. A kernel smoother such as lowess is simple and efficient, works well with binned data [26], and is computationally tractable.

After smoothing, we apply a peak detection algorithm to both re-watching session counts and play event counts. The algorithm we use is a variant of the TwitInfo [19] algorithm. It uses a weighted moving average and variance to detect unusually large number of events in time-series data, which applies well to the video context. We tested with different parameters in the algorithm to fit the time scale of our analysis, which is much shorter (the order of seconds and minutes) than what TwitInfo dealt with (hours and days).

One reason for using both replay and play counts is that they might capture different behaviors. We observe that video content includes both a time-specific event (e.g., a visual transition from a talking head to a slide) and a coherent segment that spans a longer period of time (e.g., a two-minute long explanation of a theorem). Play events capture a more precise

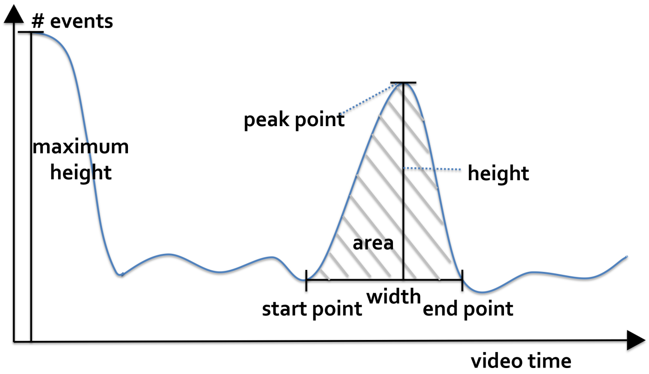


Figure 4. The location of a peak is determined by three time points (start, peak, and end). Width, height, and area determine the shape, sharpness, and intensity of the peak.

timing of an event in a video, generally resulting in sharper, spiky peaks. They respond better to student activities at one-second granularity. Re-watching session counts tend to capture segments that occur over a longer period of time better, generally resulting in smoother, wider peaks.

When a re-watching session peak and a play event peak overlap, we note that they point to a single event. When two peak windows overlap, we pick the replay peak because replay counts are always higher than play counts, possibly resulting in more informed peaks.

The features of a peak, such as width, height, and area, can indicate the strength of students' collective, time-specific interest. We compare these features between video types and student contexts. Previous work considered similar constructs in modeling temporal profiles of search queries [12]. A peak is characterized by descriptive properties as shown in Figure 4. It includes both start and end time markers, which determine the width or time duration of a peak. The peak point is the highest point between the [start, end] range, which determines the height. Finally, the area under a peak is the sum of event counts during the peak time window, which denotes the relative significance of a peak against the entire video. Multiple peaks of differing profiles might appear within a video clip. In reporting height, width, and area, we normalize the values by scaling between 0 and 1 to address high variability in event counts and durations across videos. For width, height, and area, we take a normalized range against the video duration, the maximum number of events, and the sum of all event counts, respectively.

Peak Profile Comparison

We now explore peak profiles for different video styles and watching behaviors. Overall, the mean number of peaks in a video was 3.7 (std=2.1). Of those, 2.2 (std=1.8) were replay peaks, and 2.3 (std=1.5) of them were play event peaks, which includes 0.8 duplicate peaks per video (i.e., play and replay peaks were overlapping). Considering that a mean video length was 7.8 minutes, a peak is detected roughly every two minutes in a video. Some videos exhibited as many as 11 peaks, while others did not show a notable peak. Table 2 summarizes the results in this section.

The mean width of a peak was 2.7% (std=3.5), and the median width was 9 seconds. This means that peaks in our analysis generally spanned less than 10 seconds including the rise and fall, which can point to highly time-specific events in a video. In the next section we attempt to explain what kind of events might be responsible for a peak.

The mean of normalized peak height was 7.7% (std=10.4) of the maximum height. This indicates that most peaks were quite small when compared against the maximum value of the measure. For play events, the maximum height was autoplay events at the beginning of the video, which gives a practical, comparative measure of the intensity of a peak. For example, if 10,000 students watched a lecture video and a peak had a height of 50%, this indicates that 5,000 more play button clicks were made within the peak range than in the time span just before and after the peak.

Finally, the mean of normalized peak area was 4.1% (std=4.5). This value maps to the activity dominance of a peak. A dominant single peak for a video might indicate that the peak was the single most important point of interest in the video. Conversely, a video with more peaks leaves relatively smaller area for individual peaks.

lectures vs tutorials

Tutorial videos generated stronger and more numerous peaks than lecture videos. The mean number of peaks in tutorial videos was 4.1 (std=1.9), compared to 3.6 (std=2.0) in lecture videos. A Mann-Whitney's U test shows a significant effect ($Z = -2.6$, $p < 0.01$, $r = 0.09$). Furthermore, peaks in tutorial videos were wider in width ($Z = -3.1$, $p < 0.001$, $r = 0.06$), taller in height ($Z = -7.5$, $p < 0.001$, $r = 0.13$), and larger in area ($Z = -5.5$, $p < 0.001$, $r = 0.10$) than those in lectures. Where does this difference come from?

Tutorial videos generally contain step-by-step instructions about solving a problem or using a tool. Many students follow along instructions from a tutorial at their own pace, and peaks normally occur at the step boundary. For example, a statistics course included a tutorial video on running a t-test using a statistics software package. In many cases, peaks occurred when the instructor issued commands in the tool or explained a key step in the solution, which might indicate that students re-watched these steps to make sure they follow the steps correctly. On the other hand, lecture videos are less segmented in structure with more continuous flows. Our observations show that peaks in lecture videos often relate to visual transitions in the video, such as from a slide to a talking head, or explanations of important concepts, such as introducing a theorem. While these points of interest in lecture videos attract many students to re-watch, the interaction peaks are not as sharp as in tutorial videos.

first-timers vs re-watchers

Re-watching sessions generated stronger and more numerous peaks than first-time sessions. The mean number of peaks in re-watching sessions was 2.2 (std=1.7), whereas the mean was only 1.0 (std=1.3) in first-time sessions. A Mann-Whitney's U test shows a significant effect ($Z = -14.7$, $p < 0.001$, $r = 0.35$). Furthermore, re-watching session peaks

Video Group	All	Lecture	Tutorial	First timers	Re-watchers
Peaks per Video	3.7	3.6	4.1	2.2	1.0
Normalized Height	7.7%	7.1%	10.2%	1.5%	3.1%
Normalized Width	2.7%	2.6%	3.1%	3.2%	3.7%
Normalized Area	4.1%	3.9%	4.8%	4.1%	4.7%

Table 2. Peak profile comparison reporting average values across all peaks detected for each video group. Tutorial videos resulted in more peaks than lecture videos. Likewise, re-watching sessions resulted in more peaks than first-time sessions. All differences between lecture and tutorial, and first time and re-watcing were statistically significant.

Peak Category	All	Lec.	Tut.
Type 1. beginning of new material	25%	30%	12%
Type 2. returning to content	23%	25%	15%
Type 3. tutorial step	7%	0%	30%
Type 4. replaying a segment	6%	7%	1%
Type 5. non-visual explanation	39%	38%	42%
Number of videos	80	61	19
Peaks per video	3.6	3.6	3.5

Table 3. Five student activity types that lead to a peak are shown, along with their frequency distribution as manually labeled by the authors. We sampled 80 videos and labeled each peak to one of the activity types. Only Type 5 does not involve a visual transition.

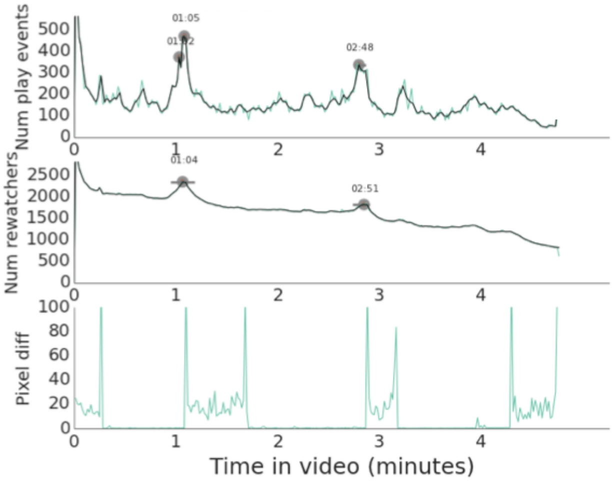


Figure 5. We visualize three streams of data to analyze interaction peaks in MOOC videos. The top graph shows play events, the middle graph shows re-watching sessions, and the bottom graph shows pixel differences over time. Detected peaks are marked with a gray point. In this example, the detected peaks coincide with a spike in pixel differences, which indicate a visual transition in video.

were wider in width (Z = -3.9, p <0.001, r = 0.07), taller in height (Z = -23.8, p <0.001, r = 0.45), and larger in area (Z = -2.9, p <0.001, r = 0.05) than first-time ones.

First-time watchers might watch videos more sequentially, because they want to master the material by watching through the lecture before diving deeper into specific parts. When re-watching, students tend to watch videos more selectively. It is notable that differences in peak height show a much higher effect size than differences in width and area. This suggests that students selectively pick parts to re-watch rather than watch through sequentially.

ANALYSIS 3. FIVE CAUSES FOR PEAKS

The peak profile analysis explains what peaks look like and how frequently they occur in different videos, but it does not reveal *why* they occur. We introduce a categorization of student activities surrounding a peak, by combining the peak profile analysis with visual content analysis. While our categorization is not conclusive, it provides an explanation of which semantic and contextual aspects of video might be responsible for a peak. This analysis suggests that no one reason can explain all peaks, and that video instructors should respond to each peak differently.

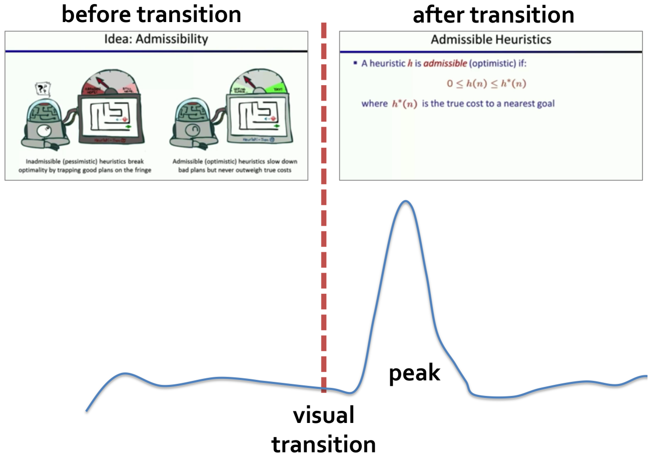
Our informal observations suggest that **visual transitions** in the video are often associated with a peak. A visual transition is a change between presentation styles shown in a video. Presentation styles in our video set are slide, code, talking head, classroom view, studio view, Khan-style tablet, and demo videos. Example transitions include changes from a slide to a talking head, a code editor to a demo video, a lecture podium view to a slide, etc. These transitions are often added at the production stage by video engineers, who mostly rely on their experiences to determine transition points. Our definition of visual transitions does not include incremental changes within a single style, e.g., an instructor typing in a new line of code in the code editor, adding an underline to highlight text, and walking a few steps in a classroom view.

Method

To explore the connection between visual transitions and interaction peaks, we apply a visual analysis technique to complement the log analysis. We use an image similarity metric that computes pixel differences between two adjacent frames to quantify the amount of visual changes in the video. Our pipeline first samples a video frame every second, computes the image similarity using the standard technique, Manhattan distance, and finally stores the pixel distance value. We visualize this data to aid the following categorization process.

We sampled 80 videos out of 862 (9.3%) while keeping the balance between video lengths, lectures vs tutorials, and production styles. This set included 20 videos from each course.

The categorization process involved two phases. In the first phase, researchers watched the selected videos, especially paying attention to the detected peaks. The goal was to construct a set of categories for peaks, using the open card sorting method [22]. As the researchers watched videos, they grouped peaks into rough categories based on common properties, such as the existence of visual transitions before or after a peak window. They discovered five groups in this gen-



Idea: Admissibility

Inadmissible (pessimistic) heuristics break optimality by trapping good plans on the fringe

Admissible (optimistic) heuristics slow down bad plans but never outweigh true costs

Admissible Heuristics

- A heuristic h is *admissible* (optimistic) if:

$$0 \leq h(n) \leq h^*(n)$$

where $h^*(n)$ is the true cost to a nearest goal

Figure 6. This peak represents the start of a new concept. The instructor started presenting a formal definition of a concept (admissibility) after changing the slide. The peak occurred when this concept explanation started.

erative process and named each. Three data streams were visualized to help with the categorization process, namely play events (Figure 5 top), re-watching sessions (Figure 5 middle), and the pixel differences (Figure 5 bottom). In the second phase, a researcher labeled all peaks in the 80 videos to one of the categories generated in the first phase.

Results

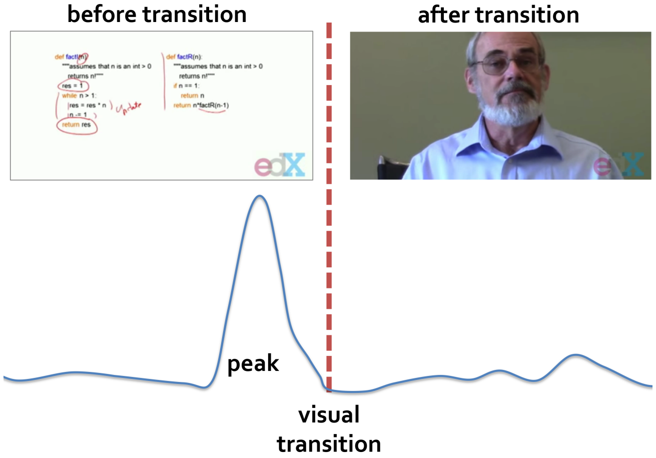
Overall, 61% of the categorized peaks involved a visual transition before, and/or after the peak. The categories, their descriptions, and frequency are shown in Table 3. We now describe each student activity category in detail.

Type 1: starting from the beginning of a new material

In this category (25% of all peaks), students browse to the beginning of a new material, such as a new concept, example, or theorem. A peak caused by such activity includes a visual transition that precedes the peak. This indicates that students are interested in the content that comes after the visual transition, which is often where new units start. Students might want to review a confusing concept after mastering earlier ones, or re-visit a theorem proof sequence. These peaks might indicate good points to cut the longer video into shorter segments, because they correspond to the beginning of a semantically different unit. Figure 6 shows an example from an AI course where a formal description of a concept (admissibility) started after presenting a motivating idea.

Type 2: returning to missed content

In this category (23% of all peaks), students return to visual content that disappears shortly after. A peak caused by such activity includes a visual transition that follows shortly after the peak. Often, the content that disappears is slides, code snippets, or board notes, but not talking heads or zoomed out views. An interpretation is that there is a pacing issue in the video. The visual transition was maybe too abrupt, not giving enough time for students to fully digest the content that disappeared. They need more time on the material, but the video view suddenly changed and prevented access to the material.



```
def factI(n):
    """assumes that n is an int > 0
       returns n!"""
    res = 1
    while n > 1:
        res = res * n
        n -= 1
    return res
```

Update

```
def factR(n):
    """assumes that n is an int > 0
       returns n!"""
    if n == 1:
        return n
    return n*factR(n-1)
```

Figure 7. This peak represents students returning to see the code snippet slide that disappeared after transitioning into the talking head. An abrupt transition might not give students enough time to comprehend what's presented.

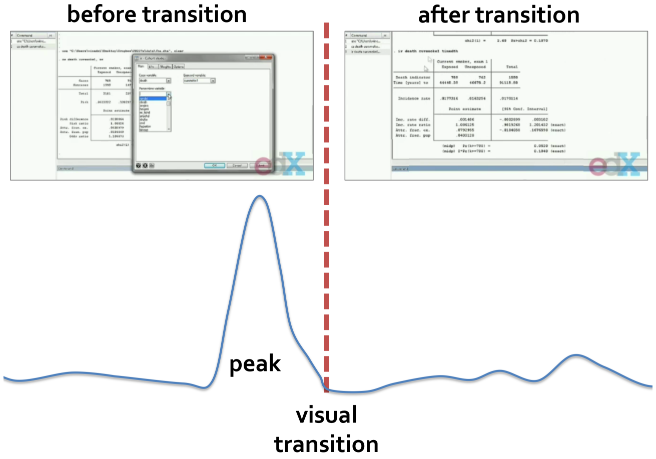
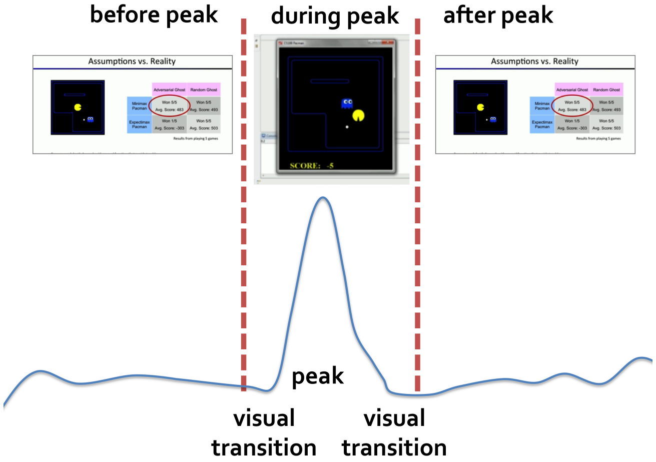


Figure 8. This peak represents students returning to a procedural step demonstrating how to run a command inside a statistics package. Students are more interested in following along the steps than the result afterward, probably because they can see the same result in their own application as well.

Also, note that what is shown during this peak type is often the final content that is complete, such as fully working code or a complete bullet point list. Many instructors make slides that advance progressively instead of showing everything at once to keep students' attention focused. When re-watching, students might want to skip to the final result without repeating all intermediate steps. Figure 7 shows an example where the code snippet suddenly disappeared and transitioned into the instructor talking.

Type 3. following a tutorial step

This category (7% of all peaks) is students following steps in the tutorial. Tutorials often contain step-by-step instructions students can follow, in the form of issuing a command or selecting a menu item from an application. Many students pause or replay right before an action takes place, possibly trying to replicate the step in their own tool. Since this was



Assumptions vs. Reality

	Adversarial Ghost	Random Ghost
Minimax Pacman	Won 5/5 Avg. Score: 483	Won 5/5 Avg. Score: 493
Expectimax Pacman	Won 1/5 Avg. Score: -303	Won 5/5 Avg. Score: 503

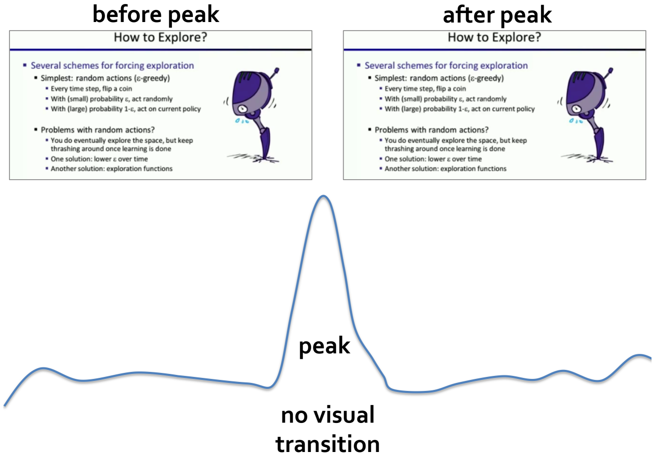
Results from playing 5 games

Assumptions vs. Reality

	Adversarial Ghost	Random Ghost
Minimax Pacman	Won 5/5 Avg. Score: 483	Won 5/5 Avg. Score: 493
Expectimax Pacman	Won 1/5 Avg. Score: -303	Won 5/5 Avg. Score: 503

Results from playing 5 games

Figure 9. This peak represents a short range of interesting segment surrounded by visual transitions before and after. The instructor launched a game application that demonstrates the concept discussed. This engaging demo might have encouraged students to return to it.



How to Explore?

- Several schemes for forcing exploration
  - Simplest: random actions (ε-greedy)
    - Every time step, flip a coin
    - With (small) probability ε, act randomly
    - With (large) probability 1-ε, act on current policy
  - Problems with random actions?
    - You do eventually explore the space, but keep thrashing around once learning is done
    - One solution: lower ε over time
    - Another solution: exploration functions

How to Explore?

- Several schemes for forcing exploration
  - Simplest: random actions (ε-greedy)
    - Every time step, flip a coin
    - With (small) probability ε, act randomly
    - With (large) probability 1-ε, act on current policy
  - Problems with random actions?
    - You do eventually explore the space, but keep thrashing around once learning is done
    - One solution: lower ε over time
    - Another solution: exploration functions

Figure 10. This peak represents important remarks from an instructor, without any visual transitions in the video. In this example the instructor was making an important point about random actions in reinforcement learning, the key topic of this AI lecture.

a recurring pattern in many of the tutorial videos, we assign a separate category. Figure 8 shows an example from a tutorial video where the instructor in the statistics course demonstrated how to run a command from a statistics package.

Type 4. replaying a brief segment

In this category (6% of all peaks), visual transitions are located both before and after the peak. This indicates that students are interested in the content within the peak range. While much less common than the other types, this type gives more specific information about student behavior because reasons explaining both Type 1 and 2 can be applied here. Figure 9 shows an example where the instructor briefly showed a demo application (during peak), and explained an underlying concept before and after the demo.

Type 5. repeating a non-visual explanation

In this category (39% of all peaks), students return to parts of a video that have no visual transitions nearby. What triggers a peak is non-visual activities in the video, such as a verbal instruction with semantic importance. We note that in many cases these peaks represent instructors introducing an important concept, re-emphasizing what has already been covered visually, or making a joke that results in a burst of laughter. Figure 10 shows an example where a peak occurred within a single slide. Here the instructor of the AI course explained the concept of taking random actions to force exploration in reinforcement learning, which was the main topic of the video.

Are there differences between peak types?

We compared normalized width, height, and area between peak types to see if peak categories, defined by the semantics of the video, map to differences in the peak profile. We first compared peaks accompanying visual transitions (Type 1, 2, 3, 4) and peaks with non-visual explanation (Type 5). A Mann-Whitney's U test shows a significant effect of height ($Z = -3.0$, $p < 0.01$, $r = 0.18$) and area ($Z = -1.9$, $p < 0.05$, $r = 0.11$), but not of width. This shows that peaks were taller and larger in size when they had visual transitions nearby. One explanation might be that visual transitions, occurring at the exact same time for all students, lead students to act similarly around them. On the other hand, start and end times of a salient activity are less clear for non-visual explanations.

Next, we looked at differences between individual categories. A Kruskal Wallis test revealed a significant effect of category on normalized height ($\chi^2(4)=19.6$, $p < 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed the significant differences between Type 1 and Type 3 ($p < 0.01$, $r = 0.33$), and between Type 3 and Type 5 ($p < 0.001$, $r = 0.32$). This suggests that tutorial step peaks (Type 3) were significantly taller than new material peaks (Type 1) or non-visual explanation peaks (Type 5). There was no significant effect found for normalized width or area. One explanation might be that tutorial steps have a clear timestamp and span a shorter period of time. For example, time between a tutorial instructor entering a command and hitting enter can be very short. The student needs to pause the video within a very short time range to capture the timing with the full command entered. For new materials and non-visual explanations, a few seconds of difference is not crucial, which might lead to smoother peaks.

DESIGN IMPLICATIONS FOR MOOC VIDEO INTERFACES

The micro-level analysis of students' video interaction introduced in this paper can guide the design of better video learning experiences. Our analysis shows that students interact with MOOC videos differently, depending on the visual, pedagogical, and stylistic properties of the video. A primary finding from both the dropout and peak analyses is that students selectively pick parts of videos to watch. And the parts they choose tend to converge to form peaks. We argue that course instructors, video production engineers, platform designers, and even students can benefit from such information. We present a set of design implications from our results for different types of learners and videos addressed in this paper.

[authoring] **Avoid abrupt visual transitions.** Type 2 peaks are likely to indicate too fast or abrupt transitions. These peaks often accompany informative slides, which can be made available outside the video as a screenshot or thumbnail for easier scanning and reviewing. Excessive visual transitions should be avoided because they might prevent students from referring to earlier content.

[authoring] **Make shorter videos.** Long lecture videos lead to a higher dropout rate. When determining points to segment long videos, Type 1 peaks can be useful points because students watch from the beginning of that segment.

[interface] **Enable one-click access for steps in tutorial videos.** Important steps in a tutorial get clear peaks. These peaks can be used to automatically mark steps in a video, making it easy for students to non-sequentially access these points without having to rely on imprecise scrubbing. Tutorial video interfaces such as ToolScape [14] adds an interactive timeline below a video to allow step-by-step navigation.

[interface] **Provide interactive links and screenshots for highlights.** Type 2 peaks suggest that missing content forces students to return. Providing static screenshots of the peak-creating informative frames might reduce the navigation overhead for students. Video interfaces might even consider multi-track streams, showing slide and instructor in separate channels that are available all the time. Type 5 peaks attract students with non-visual information, and our observation suggests that instructors make important points in these peaks. Interactive links to these points can be useful for students willing to find them later, which is especially difficult due to the lack of visual cues.

[interface] **Consider video summarization for selective watchers.** A common interaction pattern in our results is non-sequential and selective watching. Students re-watching videos tend to non-sequentially seek their points of interest. Peaks can be used to effectively summarize highlights from a video, which can be useful for students who re-watch or skim through the content while auditing.

MOOC video analytics platform

Techniques presented in this paper can provide stakeholders in a MOOC with richer data about micro-level video interaction, which can help them make data-driven decisions about planning, recording, editing, and revising videos. To support exploration of in-video interaction data, we are currently building a prototype MOOC video analytics platform. In addition to showing basic statistics per video, the enhanced video player synchronizes the video playhead with an overlay time bar on the visualization (Figure 11). This interface enables visually connecting deep-linked video content to points with salient patterns in the graph. We expect to support the sensemaking process for course instructors, video production engineers, and platform designers.

Course instructors can use MOOC video analytics to respond to students' interest and confusion while a course is being offered. Further, they can also use data-driven metrics to revise videos for the next offering of the course. **Video production engineers** can better allocate their resources in the production effort. One concrete use case is to avoid excessive visual transitions that lead to Type 2 peaks. **Platform designers** can benefit from MOOC video analytics to enhance the video player interface. For example, they can attach interactive bookmarks for peaks to improve in-video navigation.

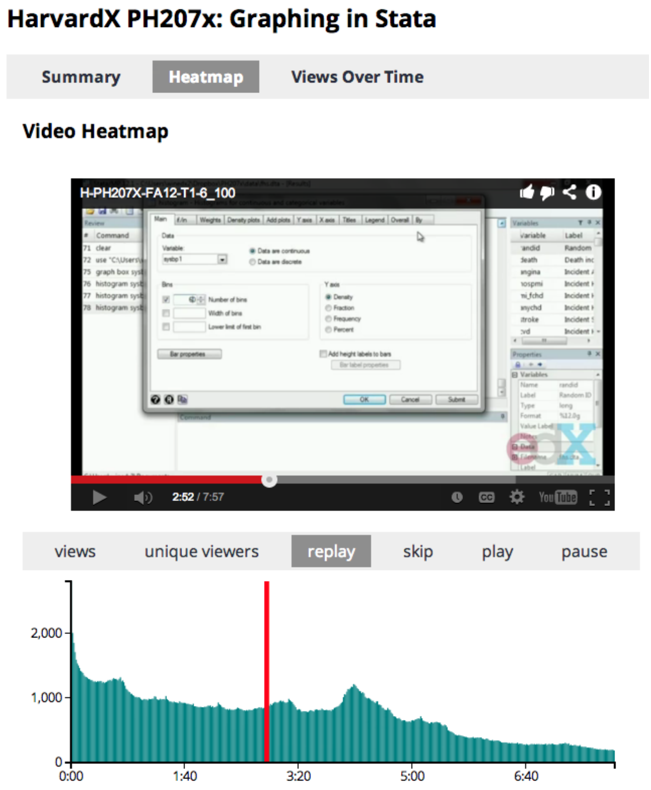


Figure 11. Our prototype video analytics dashboard supports synchronized video playback for various interaction measures.

While the analysis for this paper was done offline after the courses were complete, the analytics platform can also handle streaming events. This allows running our analytics framework for currently active courses, so that instructors can address student confusion inferred from the streaming video analytics during virtual office hours or in discussion forums.

LIMITATIONS

While our analysis methods identified video navigation patterns, understanding *why* we see these patterns is difficult. Because MOOCs do not have access to a broader learning context of a student, log entries cannot accurately represent learners' real intent (e.g., play a video but not watch). Also, video interactions might depend on other pedagogical methods in a MOOC such as problem sets, discussion forums, and exams. Furthermore, presentation quality or storyline might also affect which parts of the video students come back to watch, but our analysis does not incorporate such data. Finally, our analysis does not consider different learner goals in MOOCs, such as completing, auditing, and disengaging [15]. Per-group analysis of our techniques might reduce noise and help us better reason about the dropout and peak results.

FUTURE WORK AND CONCLUSION

This paper provides an in-depth look into how students interact with MOOC videos. We analyze data from four live courses on edX, focusing on in-video dropout rates, interaction peak profiles, and student activity categorization around peaks. We believe our data-driven analytic methods can help improve the video learning experience.

For future work, we plan to analyze more courses, data streams, and interaction patterns. We hope to analyze humanities and professional courses, and compare results against the current data from science and engineering courses. Another potential data stream is text from transcripts, textbooks, and lecture slides. Text analysis can complement vision-based techniques. In contrast to peaks, dips in viewership and interaction counts might be an informative pattern to investigate. Dips might represent boredom and loss of interest.

ACKNOWLEDGMENTS

This work was funded in part by Quanta Computer. Juho Kim is supported by the Samsung Fellowship.

REFERENCES

1. Bostock, M., Ogievetsky, V., and Heer, J. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (2011), 2301–2309.
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. Carlier, A., Charvillat, V., Ooi, W. T., Grigoras, R., and Morin, G. Crowdsourced automatic zoom and scroll for video retargeting. In *Multimedia '10*, ACM (2010), 201–210.
4. Chorianopoulos, K. Collective intelligence within web video. *Human-centric Computing and Information Sciences 3*, 1 (2013), 10.
5. Cleveland, W. S. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician 35*, 1 (1981), 54–54.
6. edX. edX Insights. `https://github.com/edx/insights`.
7. edX. Tracking Logs – edX 0.1 documentation. `http://data.edx.org/en/latest/internal_data_formats/tracking_logs.html`.
8. Girgensohn, A., and Boreczky, J. Time-constrained keyframe selection technique. In *Multimedia Computing and Systems*, vol. 1 (1999), 756–761 vol.1.
9. Google. YouTube Analytics. `http://www.youtube.com/yt/playbook/yt-analytics.html#details`.
10. Guo, P. J., Kim, J., and Rubin, R. How video production affects student engagement: An empirical study of mooc videos. In *Learning at Scale 2014, to appear* (2014).
11. Hou, X., and Zhang, L. Saliency detection: A spectral residual approach. In *CVPR '07* (2007), 1–8.
12. Jones, R., and Diaz, F. Temporal profiles of queries. *ACM Transactions on Information Systems (TOIS) 25*, 3 (2007), 14.
13. Kairam, S., Morris, M., Teevan, J., Liebling, D., and Dumais, S. Towards supporting search over trending events with social media. In *ICWSM '13* (2013).
14. Kim, J., Nguyen, P., Weir, S., Guo, P., Gajos, K., and Miller, R. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *CHI '14, to appear*, ACM (2014).
15. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *LAK '13*, ACM (2013), 170–179.
16. Kleinberg, J. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery 7*, 4 (2003), 373–397.
17. Kulkarni, A., Teevan, J., Svore, K. M., and Dumais, S. T. Understanding temporal query dynamics. In *WSDM '11*, ACM (2011), 167–176.
18. Li, F. C., Gupta, A., Sanocki, E., He, L.-w., and Rui, Y. Browsing digital video. In *CHI '00*, ACM (2000), 169–176.
19. Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., and Miller, R. C. Twitinfo: aggregating and visualizing microblogs for event exploration. In *CHI '11*, ACM (2011), 227–236.
20. Olsen, D. R., and Moon, B. Video summarization based on user interaction. In *EuroITV '11*, ACM (2011), 115–122.
21. Risko, E., Foulsham, T., Dawson, S., and Kingstone, A. The collaborative lecture annotation system (clas): A new tool for distributed learning. *Learning Technologies, IEEE Transactions on 6*, 1 (2013), 4–13.
22. Rugg, G., and McGeorge, P. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems 14*, 2 (1997), 80–93.
23. 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, to appear* (2014).
24. Shaw, R., and Davis, M. Toward emergent representations for video. In *Multimedia '05*, ACM (2005), 431–434.
25. Smoliar, S., and Zhang, H. Content based video indexing and retrieval. *MultiMedia, IEEE 1*, 2 (1994), 62–72.
26. Wand, M. Fast computation of multivariate kernel estimators. *Journal of Computational and Graphical Statistics 3*, 4 (1994), 433–445.
27. Wickham, H. Bin-summarise-smooth: a framework for visualising large data. Tech. rep., had.co.nz, 2013.
28. Wistia. Wistia Product Features. `http://wistia.com/product#analyze`.
29. Yew, J., Shamma, D. A., and Churchill, E. F. Knowing funny: genre perception and categorization in social video sharing. In *CHI '11*, ACM (2011), 297–306.