Flagpoll: A Video Annotation System to Improve Learning

Mathieu Rolfo
Stanford University
rolfo@cs.stanford.edu

Gavin Nelson
Stanford University
gmn@stanford.edu

Jonathan Leung
Stanford University
jhmleung@stanford.edu

ABSTRACT
Massive open online courses (MOOCs) have the potential to revolutionize education with technology, but as of now the actual lecture viewing process remains mostly unimproved. While there are a limited number of techniques that improve the MOOC viewing experience like in-video quizzes, we believe there is room in the design space for alternatives. In this paper, we propose Flagpoll - a Chrome extension that allows learners to flag video timestamps with labels, and view other learners' labels. This leverages principles of hands-on learning and social learning from the education literature[8]. We tested three user groups; passive lecture viewing, lecture viewing with the ability to flag, and lecture viewing with the ability to flag as well as view previous user flags. The experiment conducted did not find a statistically significant difference in the means between any of our three groups (one-way ANOVA, ($F(2, 27)=2.1$, $p=0.148$)). We then discuss explanations for our results, and potential avenues for further research.

Author Keywords
MOOC; Online Learning; Student Engagement; Video flagging

INTRODUCTION
The Internet has enabled one of the most prominent changes to education to date: virtual learning platforms. In-person learning has become more virtually situated; digital forums for question-answering, flipped classrooms, and recorded video lectures are hybrid platforms as well as prominent examples of this trend. Additionally, MOOCs (massive open online courses) have emerged as a very prominent platform for entirely virtual learning, with the capability to scale far beyond traditional learning environments[1].

This new environment is a highly effective way to provide education to millions of learners. However, at such large scale, it is important to optimize the educational process as much as possible, and the lecture viewing experience is one area ripe for improvement[3][4]. In-video dropout rates are high [6], and video lectures lose out on the engagement and interactivity of in-person lectures. Education literature has suggested that "hands-on learning" is better for student engagement[7] than passive viewing.

There is one primary technique used in MOOCS to ameliorate this problem: in-video quizzes. In-video quizzes are placed in junctions within the video, usually between topic transitions, to ensure that a viewer understood what was previously discussed before moving on. In-video quizzes provide a level of engagement while ensuring a student's understanding[5]. Furthermore, studies have shown that almost 75% of students engage in these quizzes even without instructor prompting, implying that learners find them useful[5]. However, they do have some drawbacks. They require additional effort from professors to create quizzes and implement them in the video, which limits the ubiquity of in-video quizzes. Additionally, the method hinges on the quality of the questions: a poorly written question could distract learners without providing any benefits.

We suggest another potential approach, building on the concept of annotation as developed in other studies[11][2]. In Zyto et al.'s study, they provide evidence that document margin annotating of online reading improved learning outcomes for students. When students annotate document margins, they interact more deeply with the content, leading to increased understanding[11]. One benefit to the annotation approach is that students can set their own effort levels to meta-lecture activities, and can potentially provide benefits with less effort on students' parts. Additionally, it does not require any effort from instructors to operate, which could potentially allow annotation-based techniques to generalize to non-lecture domains. In industry, Soundcloud, YouTube, and Waze (a navigation application) allow for user annotations[9].

There is also a large amount of evidence from the education literature that the social element of learning is important[8]. MOOCs by their virtual nature can obscure that many students are learning together. While students can communicate with others on forums, they lose the ability to get feedback about other students during the lecture-viewing experience. Similarly, we feel that students can benefit from knowing information about the lecture before viewing, and can adjust their attention levels accordingly. We have seen limited evidence of methods like this in the existing research, and explore this here.

We propose Flagpoll, a browser extension overlaying videos that allows viewers to flag sections of the video with labels as they watch. Additionally, the extension displays a timeline showing previous viewers' flags of the video, serving as a poll of other users' opinions of the difficulty and engagingness of a given lecture. This design reflects our underlying two hypotheses, which we evaluate in this study:

1. Students who view lectures while using an annotation tool to flag sections, due to hands-on learning effects, will have increased information retention over students who watch lectures passively.
2. Students who view lectures while using an annotation tool to flag sections, that also displays the flags of other users' about the video, will have increased information retention compared to passive viewers and flagging-only users. This is because the other users' flags will identify points of interest, which viewers will then pay more attention to while viewing.

METHOD AND EXPERIMENT

Flagpoll

Flagpoll is a Google Chrome extension that creates a flagging interface and timeline view below YouTube videos. Our interface provides two buttons, Difficult and Uninteresting, which mark the timestamp of the video whenever clicked. Users can also click on created timestamps to remove them.

We chose Difficult and Uninteresting for a number of reasons. First, when considering the difficulty of a lecture, we wanted users to focus on parts they struggled with the most. Second, when thinking about user disengagement, finding the most uninteresting sections would allow professors to take note of unnecessary parts of lectures and edit their lectures [10].

These buttons sat below any Youtube video in a small standalone box. Finally, easy or medium difficulty sections do not provide much information, only that a user felt they did not need to focus more when we wanted to focus on sections of increased or complete lack of focus.

For the timeline, we incorporated a third party library, Vis.js, to display flag times to users. This timeline was displayed below the Difficult and Uninteresting buttons, taking up the rest of the page for our extension when viewed in default mode **more about the extension**

Experiment

We recruited 30 participants (17 male, 13 female) and divided them into equally-sized groups of 10 for our three conditions. The participants ages ranged from 18 to 66 and they all have some degree of a college education, and were recruited from mailing lists.

Each subject participated in a study that took approximately 30 minutes. They were directed to an 18 minute segment of a lecture titled, "Intro to Crypto and Cryptocurrencies," a Bitcoin and Cryptocurrency online lecture series from Princeton University. Preliminary tests conducted showed that a nontechnical video did not provide enough difference in test scores.

The cryptocurrencies lecture was chosen because it provided a conceptually non-trivial topic that did not depend on mathematics or other prior knowledge. Participants were also screened for prior knowledge. Preliminary research on the video for the study suggested that TED talks were too comprehensible to get low scores on quizzes that focused on conceptual understanding.

Participants then were directed to a sample of SAT math questions online and were instructed to spend five minutes completing the problems. These SAT questions were designed to distract subjects from remembering video content, simulating loss of information over time. After the five minutes expired, subjects were given a Google forms quiz with nine questions relating to the cryptocurrency lecture video.

Initially a control group of ten participants participated in this study without the implementation or use of the Flagpoll plugin. They watched the video with a standard YouTube interface and then completed both subsequent quizzes. We hoped to see our scores increase as we introduced different aspects of the plugin.

A second group watched the video with only the Flagpoll interface for flagging difficult and uninteresting time stamps. This group was asked to do the same study but to flag at least one "difficult" and one "uninteresting" moment. The flagged data was collected after they watched the video and they then completed both quizzes. We then aggregated this data to fill our timeline, giving us an idea of what subjects found interesting or difficult.

A third group watched the video with the full Flagpoll implementation. They had the same flagging options but could also see the time line of marked flags from the second study group's data. After watching and flagging the video, the third group completed both quizzes.

This study was conducted with only these three groups, rather than with the fourth group of timeline-only. A control group was necessary for results. The second study where participants just flagged provided data on how the active flagging improved quiz scores. The third study was conducted to convey the use of the users input to future participants. If users see crowdsourced flags as they are actively flagging videos, we felt it would provide more of a collaborative learning mentality to the participants. We believed that the hands-on aspect of the tool was fundamental, and so did not believe that the timeline-only trial would show differences from control. This also let us increase the number of participants on our other conditions.

The experiment was set up by trimming the cryptocurrency video to the desired 18 minute segment and rehosting it on YouTube. The nine question quiz was hosted on Google Forms, and the data collection was conducted in Google Sheets. Each participant was invited to take the quiz which contained details of the steps. The video quiz contained multiple choice and check-the-box questions ranging from discussing the steps of certain hash functions to describing the properties of hashes, to minimize the success of guesses.

RESULTS

Data

The table below shows some summary statistics for our data. On our quiz, out of 9 points, the overall mean amongst the

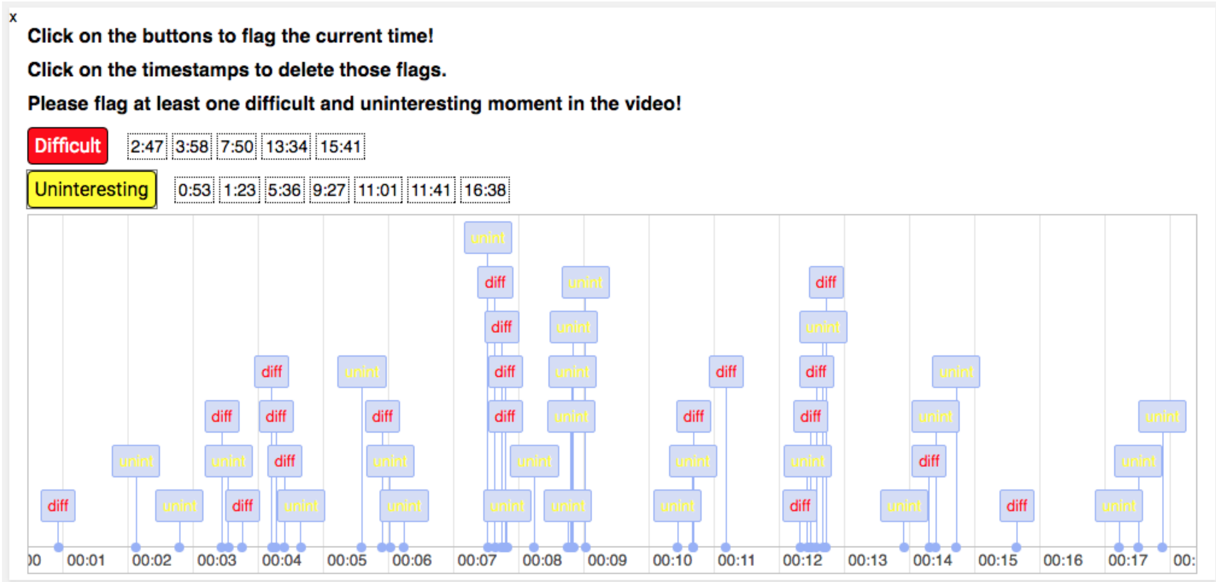
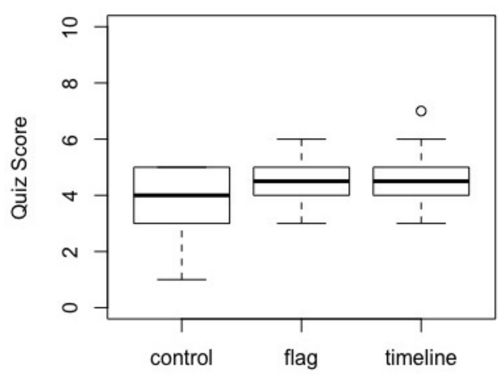


Figure 1. Example of flagging and time line environment presented

three conditions was 4.27 and our median was 4.0. The lowest score obtained was a 1, and the highest was a 7.

	Control	Flagging Only	Flag and Timeline
Mean St. Dev.	3.7 1.25	4.4 0.97	4.70 1.16

The data is represented in boxplot format below.



Analysis

As reported earlier, we found a difference in means between our control group and each of the conditions, where both flagging and flagging plus timeline groups had higher means. To analyze our data for statistical significance, we ran a one-way ANOVA (analysis of variance) test. Because we did not run trials for the full 2x2 matrix of conditions, we did not run a two-way ANOVA test on our data. There were no outliers from the data to remove.

However, our one-way ANOVA did not reveal a statistically significant difference in quiz scores amongst the three conditions ($F(2, 27)=2.1$, $p=0.148$). For further investigation, we ran pairwise t-tests without correction on our conditions, and found $p=0.15$ for control-timeline, $p=0.40$ for control and flag, and $p=0.52$ for flag and flag+timeline.

DISCUSSION

We did not find statistically significant results to support our hypotheses. One explanation for this is limited sample size - we had differences in our condition means, but the variation amongst our small samples obscured this. Collecting data from a significantly larger group of participants would have required tens of more hours of testing which was unfeasible for the scale of this project.

We also focused our efforts towards college-age learners, though we had some variation in our sample. Furthermore, our minimum age was 18, cutting off younger generations of students. If we had more time, we would have liked to test our hypotheses on different populations, such as younger children.

There might also be technical improvements that could have led to stronger results. For example, users often desired to watch the lecture in a full screen view, but could not interact with the Flagpoll UI from full screen due to our implementation. Thus, they were forced to adjust their viewing style to fit our experimental setup. Additionally, the user-generated flags were not displayed on the timeline, which made the information less accessible and less comparable to existing flags.

However, we still believe there are some ideas that can be adapted from Flagpoll to other research. For the purposes of this study, it is also difficult to tell the participants' desire to flag from this study. Each participant was instructed to flag *at least* one difficult and uninteresting section even though many flagged more, but if there were no minimum bar, as would occur in the real world, it would be interesting to see how users would behave. Additionally, we believe that giving users information about sections of the video, either about video as a whole (e.g. this lecture has been noted as particularly difficult) or sections (e.g. this is the hardest part of the lecture), would be useful for learners. This data could be derived either from user flags, or from machine learning techniques.

CONCLUSION

In this paper we present Flagpoll - a Google Chrome extension that aims to bring enhanced learning to online lecture videos through flagging difficult and uninteresting sections of video and viewing other users' flags. Flagpoll's strengths include bringing a classroom collaborative-learning mentality to online lecture viewing, collecting crowd sourced data on difficult and uninteresting sections of lectures, as well as helping students stay more focused on the video at hand. Unfortunately the study run was unable to find statistical significance disproving the null hypothesis. Future work building off of Flagpoll has the potential to impact learning nonetheless.

ACKNOWLEDGMENTS

We would like to thank Michael Bernstein, Rob Semmens, and Kesler Tanner for their feedback and guidance throughout this research process.

REFERENCES

1. Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*. ACM, 687–698.
2. Ian Glover, Zhijie Xu, and Glenn Hardaker. 2007. Online annotation âĂŞ Research and practices. *Computers and Education* 49, 4 (2007), 1308 – 1320. DOI: http://dx.doi.org/10.1016/j.compedu.2006.02.006
3. Philip J Guo, Juho Kim, and Rob Rubin. 2014. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 41–50.
4. Chen-Tai Kao, Yen-Ting Liu, and Alexander Hsu. 2014. Speeda: adaptive speed-up for lecture videos. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology*. ACM, 97–98.
5. Geza Kovacs. 2015. QuizCram: A Question-Driven Video Studying Interface. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 133–138.
6. Geza Kovacs. 2016. Effects of In-Video Quizzes on MOOC Lecture Viewing. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 31–40.
7. Heidi L. Lujan and Stephen E. DiCarlo. 2006. Too much teaching, not enough learning: what is the solution? *Advances in Physiology Education* 30, 1 (2006), 17–22. DOI: http://dx.doi.org/10.1152/advan.00061.2005
8. D.L. Schwartz, J.M. Tsang, and K.P. Blair. 2016. *The ABCs of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them*. W. W. Norton. https://books.google.com/books?id=GjOTCgAAQBAJ
9. Thiago H Silva, Pedro OS Vaz de Melo, Aline Carneiro Viana, Jussara M Almeida, Juliana Salles, and Antonio AF Loureiro. 2013. Traffic condition is more than colored lines on a map: characterization of waze alerts. In *International Conference on Social Informatics*. Springer, 309–318.
10. Mohammad Soleymani and Martha Larson. 2010. Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. (2010).
11. Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1883–1892.