Who Downloads Online Content and Why?

Katherine Brady
Department of Electrical Engineering and Computer Science
Vanderbilt University
katherine.a.brady@vanderbilt.edu

Gayathri Narasimham
Institute for Digital Learning
Vanderbilt University
gayathri.narasimham@vanderbilt.edu

Douglas Fisher
Department of Electrical Engineering and Computer Science
Vanderbilt University
douglas.h.fisher@vanderbilt.edu

ABSTRACT
Online learners sometimes prefer to download course content rather than view it on a course website. These students often miss out on interactive content. Knowing who downloads course materials, and why, can help course creators design courses that fit the needs of their students. In this paper we explore downloading behavior by looking at lecture videos in three online classes. We found that the number of days since a video was posted had the strongest relationship with downloading, and non-technical considerations, such as typical classroom size in a student's home country, matter more than technical issues, such as internet speed. Our findings suggest that more materials will be downloaded when a course will be available for limited time, students are less familiar with the language of instruction, students are used to classrooms with a high student-teacher ratio, or a student's internet speed is slow. Possible reasons for these relationships are discussed.

Author Keywords
Accessibility; MOOC; Online Education

ACM Classification Keywords
K.3.1. Computers and Education: E-Learning

INTRODUCTION
Interactive content keeps learners engaged, provides quick feedback and improves learning outcomes [14, 15]. In online classes, interactive content, such as short quizzes and instant message portals, can be embedded using web technologies, though these technologies usually don't allow the content to be downloaded. Logs of online courses indicate that many students choose to download materials when they can. Knowing when and why learners will download materials can help course creators tailor content for their audience.

In this paper, we investigate downloading by looking at the influence of video properties, time, and both technical and non-technical properties of the country a video was accessed from. As far as we know, this is the first study to look specifically at downloading behavior. Previous studies have touched on reasons a learner may choose to download, but none have looked into which of these reasons plays the largest role. Our findings indicate that **course deadlines** play the biggest role, followed by **non-technical considerations**. Technical difficulties such as slow internet had weaker, but significant correlations with downloading. Course creators can use these findings to tailor content for the viewing audience.

BACKGROUND
Past studies of online learning behavior mention downloading, though none concentrated specifically on why students download. In interviews, students have mentioned plans to download course materials to either go through at their own pace [4, 12], or watch in environments where internet is not available [24]. Demographically, learners who download videos are more likely to be from India than the United States, Russia or Spain [21]. Learners who download most of what they watch also tend to join the course later, be younger, male, and employed [11].

Most past studies of approaches to online courses concentrated on learner performance [6, 7, 12, 13, 20]; however navigation strategies [9, 16, 18], course involvement [10, 22], and motivation [11, 13, 24] have also been studied. Bandwidth speed [1, 13, 17], student-teacher ratio [9], age [9, 10], language familiarity [6, 7, 10], and deadline timing [1, 13] have been identified as significant factors in explaining differences in learner approaches.

DATA SET
For this study we used data from the log files of three Coursera courses taught in English. These courses are currently offered as on-demand courses, but this study uses earlier sessions to study the influence of deadlines. The courses covered topics in literature, medicine, and business, and were 4-8 weeks long. Each time a video was accessed, Coursera recorded whether a learner streamed or downloaded it, the time, and the learner's IP address. We mapped these IP addresses to country level statistics on bandwidth, classroom size, and language. This meant that with the exception of a few demographic variables, such as age, our data covered all variables mentioned in the literature. A list of the variables used is shown in Table 1.

We found country level statistics on bandwidth speeds from a report on the state of the internet by Akamai Technologies [2].

L@S 2018, June 26–28, 2018, London, United Kingdom

DOI: **https://doi.org/10.1145/3231644.3231699**

To estimate the English proficiency of learners in each country we used a report of the average scores on the Education First commercial-English test by country [8]. In countries where English is widely spoken, no one takes this test; thus, we assigned a score slightly higher than the highest score in the report for those countries. The World Bank has records of student-teacher ratio in grade school classrooms across the globe since 1960 [23]. Using country level statistics limits the power of our analysis, but we believe this indicates the promise of further study with more precise statistics.

METHOD

We had five a priori hypotheses based on previous findings:

i. Peak Internet speed would have the strongest correlation with download rate, followed by other bandwidth metrics

ii. Learners would be more likely to download longer videos than short videos

iii. Learners less proficient in English would prefer to download videos

iv. Learners from Africa, Asia and South America would access videos later than Europeans and North Americans

v. Learners from countries with a low student-teacher ratio would be more likely to re-watch streamed videos

Previous research has found that large online courses are predominantly taken by richer and more educated members of society [5, 6, 20], suggesting peak internet speed would be more predictive than average internet speed. Longer videos are more susceptible to internet interruptions [19], so there are more benefits to downloading when a video is longer. Subtitles are easier to add to downloaded videos [3], so learners less proficient in English would be more likely to download. The final two hypotheses are findings from previous studies. Kizilcec and Halawa surveyed learners from 20 online courses, and found that learners from Africa, Asia and South America reported more trouble with deadlines than their peers in Europe and North America [13]. Guo and Reinecke looked at four large online courses and found that the students to teacher ratio correlated with how often learners rewatched streamed videos [9]. We also included three course log variables besides video length in our analysis. These variables were easy to obtain, but had no implications from previous findings.

To test these hypotheses, we built two correlation models. The first looked at the correlations between each variable and downloading at a per-video access level. Thirty one percent of the videos were accessed from the United States; to correct for oversampling we re-ran the analysis on the 69% of accesses from other countries, and got similar results. Our second model looked at the correlation between each country's statistics and downloading by aggregating the downloading to the percentage of accesses in a country that were downloads.

RESULTS

We plotted each variable against downloading, and, as shown in Figure 1, there is a change in the relationship between

	Variable	Description
1	Days Since Opening	Days between the video being made available and a learner accessing it
2	Week of Course	Week of a course that the video was made available
3	Video Length	Length in seconds of the video
4	Popularity	Total number of students who accessed the video
5	Percent Attack Traffic	Percent of internet traffic in a country from malicious sources
6	Average Connection Speed	Average connection speed within a country in MB per second
7	Peak Connection Speed	Peak connection speed within a country in MB per second
8	Percent Above 10 MB/s	Percentage of internet users within a country whose bandwidth is more than 10 MB per second
9	Percent Above 4 MB/s	Percentage of internet users within a country whose bandwidth is more than 4 MB per second
10	English Proficiency	Average score on the Education First commercial English proficiency test within a country among voluntary test takers
11	Student-Teacher Ratio	Average number of students per elementary school classroom within a country

Table 1. The Variables Used. The first four (1-4) come from the course logs and the next seven (5-11) come from country statistics linked to learners by their IP address. Variables 5-9 are referred to as the 'bandwidth statistics' and come from the Akamai report.

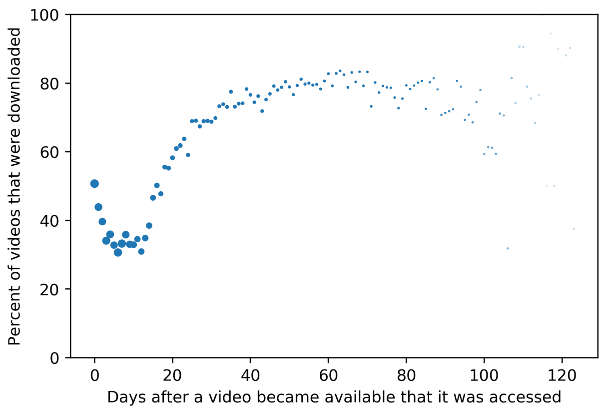

Figure 1. Percent of videos downloaded as a function of days since a video was posted. The size of each circle indicates the number of users who accessed videos on that day.

downloading and 'days since opening' after 18 days. This could be related to the assignment due dates, which were typically about three weeks after a video on which the assignment was based was released. Since the difference in downloading behavior before and after this cutoff was so strong, we partitioned data on this cutoff in the correlation models.

Variable	First 18 Days	After 18 Days
English Proficiency	-0.212	-0.055
Student-Teacher Ratio	0.212	0.032
Percent Above 10 MB/s	-0.189	-0.055
Average Connection Speed	-0.186	-0.051
Percent Above 4 MB/s	-0.185	-0.034
Popularity	-0.182	-0.173
Peak Connection Speed	-0.179	-0.047
Week of Course	0.175	0.135
Percent Attack Traffic	-0.086	-0.016
Video Length	-0.038	-0.095
Days Since Opening	-0.038	0.127

Table 2. Correlations between each variable and downloading. Variables are ordered according to the correlations in the first 18 days a video was available. All correlations are significant to $p < 0.001$

In the first 18 days that a video was available, location based attributes had strong correlations with downloading behavior, as shown in Table 2. After 18 days, variables from the course logs mattered much more. In both cases, bandwidth statistics generally had weaker correlations than 'English proficiency' and 'student-teacher ratio'.

In the country level model, we could not look at course log variables since the model was aggregated by country. None of the location based variables had a significant correlation to downloading after 18 days. In the first 18 days, 'English proficiency' had no significant relationship to downloading, but all other location based variables had strong correlations, as shown in Table 3. 'Student-teacher ratio' continued to have a stronger relationship to downloading than all but one of the bandwidth variables.

Variable	First 18 Days
Percent Above 4 MB/s	-0.47 **
Student-Teacher Ratio	0.42 **
Peak Connection Speed	-0.39 **
Average Connection Speed	-0.37 **
Percent Above 10 MB/s	-0.32 *
Percent Attack Traffic	0.30 *
English Proficiency	not significant

Table 3. Correlations between each country statistic and the percentage of videos downloaded in that country in the 18 days after each video was released. Measurements marked ** are significant at $p < 0.01$ **and those marked * are significant at** $p < 0.05$

Separate analysis found that learners who streamed videos from Africa, Asia and South America accessed videos later than their North American and European counterparts. However, when downloading accesses were included in the analysis, there was no difference in when learners accessed videos between on continents. We also found a slight positive correlation between the percent of videos that were re-watched in a country and the respective student-teacher ratios, but it was not statistically significant.

CONCLUSION AND FUTURE WORK

Of the three hypotheses regarding downloading behavior ($i. - iii.$), we found evidence to support the third (iii), that learners less proficient in English prefer to download videos. Bandwidth speeds did not have the strongest relationship with downloading in any of our models and longer videos were not more likely to be downloaded. The only supported hypothesis was that English proficiency was negatively correlated with downloading videos, though even this result was only observed at the per video access level. The strong relationship between the time when videos were made available and downloading seems to be driven by when course assignments were due. The fact that non-technical attributes play a larger role than bandwidth metrics in both models suggests the decision to download is not purely driven by technical considerations, though learners with lower bandwidth speeds were more likely to download videos.

Our findings appear to confirm the two findings from previous research ($iv. - v.$). We found a similar relationship between learners' continent and days since opening among learners who streamed videos as in Kizilcec and Halawa [13], and found a less significant relationship between student-teacher ratio and re-watching behavior than Guo and Reinecke [9].

We do not know why students from countries with larger average student-teacher ratios prefer downloading videos to streaming them, but this relationship may reflect larger cultural differences around education. Previous studies have found correlations between the student-teacher ratio and how much a society spends on education and how teacher centered the classroom is [9]. We hope to futher explore how the student-teacher ratio interacts with downloading using more precise metrics and a larger data set.

Online classrooms make education available to learners in disperate locations. However, world-wide classrooms are difficult to plan lessons for. The courses studied in this paper were designed by professors from the United States and intended for learners who streamed the videos before the assignments were due. The website logs show that many learners took a different approach from that envisioned by the course designers. We hope the results in this paper can help future course designers tailor their content.

REFERENCES

1. Aboshady, O. A., Radwan, A. E., Eltaweel, A. R., Azzam, A., Aboelnaga, A. A., Hashem, H. A., Darwish, S. Y., Salah, R., Kotb, O. N., Afifi, A. M., Noaman, A. M., Salem, D. S., and Hassouna, A. Perception and use of massive open online courses among medical students in a developing country: multicentre cross-sectional study. *BMJ Open 5*, 1 (jan 2015), e006804–e006804.
2. Akamai Technologies, and Belson, D. The State of the Internet. **https://www.akamai.com/us/en/about/our-thinking/state-of-the-internet-report/**, 2013.
3. Akbar, S. G. Online Subtitles Translator & Editor. **http://www.syedgakbar.com/projects/dst/**, 2009.
4. Belanger, Y., and Thornton, J. Bioelectricity: A quantitative approach duke universitys first mooc. Tech. rep., Duke University, 2013.
5. Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., and Emanuel, E. J. The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why? *SSRN Electronic Journal* (2013).
6. Cutrell, E., Neill, J. O., Bala, S., Nitish, B., Cross, A., Gupta, N., Kumar, V., and Thies, W. Blended Learning in Indian Colleges with Massively Empowered Classroom. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15* (2015), 47–56.
7. Dasgupta, S., and Hill, B. M. Learning to Code in Localized Programming Languages. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale* (2017), 33–39.
8. Education First. EF English Proficiency Index. **https://en.wikipedia.org/wiki/EF_English_Proficiency_Index**, 2017.
9. Guo, P. J., and Reinecke, K. Demographic Differences in How Students Navigate Through MOOCs. In *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14* (2014), 21–30.
10. Huang, J., and Manning, J. Superposter behavior in MOOC forums. In *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14* (2014), 117–126.
11. Kahan, T., Soffer, T., and Nachmias, R. International review of research in open and distributed learning. *International Review of Research in Open and Distributed Learning 18*, 6 (2017), 1–18.
12. Khalil, H., and Ebner, M. MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review. *EdMedia: World Conference on Educational Media and Technology 2014*, 1 (2014), 1305–1313.
13. Kizilcec, R. F., and Halawa, S. Attrition and Achievement Gaps in Online Learning. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15* (2015), 57–66.
14. Koedinger, K. R., Mclaughlin, E. a., Kim, J., Zhuxin Jia, J., and Bier, N. L. Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15* (2015), 111–120.
15. Kovacs, G. Effects of In-Video Quizzes on MOOC Lecture Viewing. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (2016), 31–40.
16. Liegle, J. O., and Janicki, T. N. The effect of learning styles on the navigation needs of Web-based learners. *Computers in Human Behavior 22* (2006), 885–898.
17. Liyanagunawardena, T. R., and Adams, A. The impact and reach of MOOCs: a developing countries' perspective. *eLearning Papers* (2014), 38–46.
18. Mirriahi, N., Jovanovic, J., Dawson, S., Gašević, D., and Pardo, A. Identifying engagement patterns with video annotation activities: A case study in professional development. *Australasian Journal of Educational Technology* (July 2017).
19. Mok, R. K. P., Chan, E. W. W., Luo, X., and Chang, R. K. C. Measuring the quality of experience of HTTP video streaming. *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on* (2011), 485–492.
20. Morris, N. P., Morris, N. P., Hotchkiss, S., and Swinnerton, B. Can demographic information predict MOOC learner outcomes? Can demographic information predict MOOC learner outcomes? In *Proceedings of the European MOOC Stakeholder Summit 2015* (2015), 199–207.
21. Seaton, D. T., Nesterko, S., Mullaney, T., Reich, J., and Ho, A. D. Characterizing Video Use in the Catalogue of MITx MOOCs. *Elearning Papers 37*, March (2014), 33–41.
22. Stankiewicz, A. $1 Conversational Turn Detector: Measuring How Video Conversations Affect Student Learning in Online Classes. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (2016), 81–88.
23. World Bank. Pupil-Teacher Ratio in Primary Education. **https://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS**, 2016.
24. Zheng, S., Rosson, M. B., Shih, P. C., and Carroll, J. M. Understanding Student Motivation, Behaviors, and Perceptions in MOOCs. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (2015), 1882–1895.