

Massive Open Online Courses: a New Window on Education

Daniel T. Seaton^{1*}, Yoav Bergner¹, Isaac Chuang^{1,2}, Piotr Mitros³, and David E. Pritchard¹

¹Department of Physics and Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139. ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139. ³edX and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139.

*Correspondence to: dseaton@mit.edu.

Abstract

Massive open online courses (MOOCs) offer a new window for observing student learning behavior. Time-stamped logs show when students clicked on which page of the textbook, skipped over or reviewed a section of the lecture video, asked or answered a question in the discussion forum, opened or answered a homework problem, etc. – information that is nearly impossible to obtain in a traditional on-campus course. We present an overview of such learning activities for the 108,000 participants behaved in 6.002x - Circuits and Electronics, the first course in MITx (now edX), a course whose structure mirrors a traditional on-campus course: lecture + recitation + 2 exams. We concentrate on the 7% of participants who obtained certificates because they accounted for 60% of the total time spent by all participants, and continued their activity through the entire course allowing week by week comparisons. We examine how the successful students (i.e. those receiving certificates) allocated their time amongst the various course components and study what fraction of each they accessed. This work lays the foundation for future studies of how resource use and learning habits influences learning in MOOCs.

1. Introduction

Forget the hype about MOOCs. Don't worry about the 100,000 registrants who never attempted significant homework (they had a median time in course ~ 20 minutes anyway [Seaton et. al. submitted to CACM]), or whether MOOCs reach out significantly to the disadvantaged. And most important, don't worry about whether MOOCs are a threat to colleges and their faculty – realize that they offer us a tremendous window into what we're doing as teachers, a window onto essentially all student learning activity in a course whose organization is that of an on-campus course using traditional lecture-recitation format. This paper will demonstrate that MOOCs represent a break-through way to study students' learning activity in a *traditional college course format*.

This is tremendously valuable because it has been so difficult to study much of the learning activity in an on-campus course. Do your students read their textbook? If so do they read it before class as some teachers recommend? Which resources available to on-

campus students do they actually use: their textbook, the library, other students, the various tutor rooms? Admit it, we don't know, except for some self-reported surveys.

Our data consist of the 100GB of logged student interaction data from the inaugural MITx (now edX) course, 6.002x *Circuits and Electronics*, in Spring 2012: data at least two orders of magnitude larger than analyzed in previous studies of online learning [21,10]. We restrict the study to successful students – those who earned certificates. We examine the use of different course components (e.g. lecture videos, homework, discussion forum, e-text, etc.) in terms of user time allocation and the total fraction accessed. Our results emphasize the richness of data available from a MOOC like 6.002x.

2. 6.002x, Procedures, and Data Analysis

With some modification for online delivery, the 14 weeklong units of 6.002x largely mirror a traditional on-campus course in both format and timing. The prescribed course sequence (the left navigation bar in Fig 1) comprises lecture sequences consisting of lecture videos (annotated powerpoints and actual MIT lectures) with embedded lecture questions, tutorial videos (recitation substitute), weekly homework (3-4 multi-part problems), and weekly lab assignments (interactive circuit toolbox). Overall grades were determined by homework 15%, labs 15%, a midterm 30%, and a final 40%.

Supplementary materials (top navigation bar in Fig. 1) include a course textbook (navigable page images), a TA- and student-editable wiki, and moderated student discussions. For further exploration of course structure and available resources, readers may visit the archived course (<https://6002x.mitx.mit.edu/>).

The screenshot displays the MITx course interface for 6.002x. At the top, a dark red navigation bar contains the course title 'MITx Circuits and Electronics' and several menu items: 'Courseware', 'Course Info', 'Textbook', 'Discussion', 'Wiki', and 'Profile'. Below this, a 'Courseware Index' sidebar on the left lists the course structure, including 'Overview', 'Week 1' (with sub-items like 'Administrivia and Circuit Elements', 'Circuit Analysis Toolchest', 'Basic Circuit Analysis', 'Resistor Divider', and 'Week 1 Tutorials'), and subsequent weeks up to 'Final Exam'. The main content area shows a video player for 'S1V1: Motivation for 6.002x'. The video player includes a title bar, a video thumbnail with the text '6.002x CIRCUITS AND ELECTRONICS Introduction and Lumped Circuit Abstraction', a progress bar at 0:00 / 0:53, and a speed control set to 1.0x. To the right of the video player, there is a 'SPEAKER 1' section with text describing the course as 'an extraordinarily fun course' and providing information about the curriculum and learning goals. Below the video player, there are links for 'More information given in the text.', 'Lecture Slides Handout [Clean][Annotated]', and 'Discussion: S1'. At the bottom of the page, there is a copyright notice 'Copyright © 2012. MIT. Some rights reserved.', social media icons for LinkedIn, Twitter, and Facebook, and links for 'Feedback', 'Help', and 'Log out'.

Fig. 1: Screenshot of typical student view in 6.002x. All course components are accessed from this interface. The left sidebar navigates to weekly modules containing lecture

sequences (videos and questions), homework, lab, and tutorials, while the header navigation provides access to supplementary materials: digital textbook, discussion forums, and wiki. The main frame represents the first lecture sequence; beige boxes below the header indicate lecture videos and questions.

3. Estimation of time spent on different resources

Time estimation for each participant relies on measuring the durations between a student's initial interaction with a resource and the time when they navigate away. We accumulate durations through analysis of each participant's time-series accounting for each separate course component (Homework, Book, Discussion Forums, etc.). We have evidence that those durations less than 3 seconds represent students navigating to desired resources, hence, we don't count these intervals as a legitimate student access. In addition, we add no time for durations over 1 hour, assuming that the user has disengaged from their computer much before that duration. Using alternate values of the high cutoff (20 min to 1 hr) can change overall times by 10-20%, but does not significantly alter relationships regarding time allocation among course components or total time spent by different participants.

An important point is that time accumulated is associated with the current resource. For example, if a student references the book while working on the homework, this duration is accumulated with book time. Only direct interactions with the homework are logged with homework resources. There are clearly alternatives to this approach, e.g. considering all time between opening and answering a problem as problem-solving time [21].

4. Results

4.1 Frequency of accesses

Fig. 3 shows the number of active certificate earners per day for, where large peaks occur on Sunday deadlines for homework and labs. There is a downward trend in the weeks between the midterm and the final exam (shaded regions). No homework or labs were assigned in the last two weeks before the final exam, though the weekly peaks persist suggesting that this periodicity indicates the students were otherwise occupied during workdays. We plot activity in events per active student per day for assessment-based course components and for learning-based components in Fig. 2B and 2C.

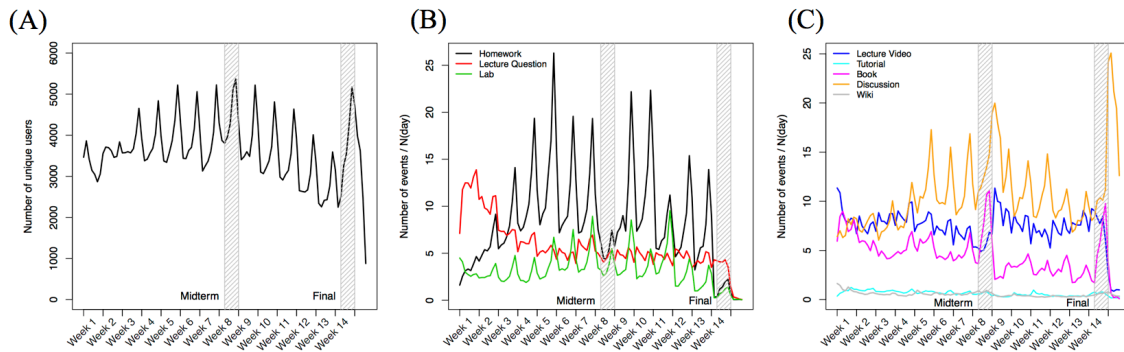


Fig. 2: From left to right, number of unique certificate earners $N(\text{day})$ active per day, their average number of accesses each day for assessment-based (middle) and learning-based course components (right panel) each day. Plot (A) highlights the periodicity and trends of overall activity. Plot (B) shows assessment-based (Homework, Lab, Lecture questions) accesses per active student. Learning-based activities (Lecture videos, Textbook, Discussion, Tutorial, Wiki) are shown in plot (C)

The activity per user is very periodic for Homework and Labs – the “for credit” assessments, as well as for the highly used discussion forum, which is the most frequently consulted resource for students doing homework. This indicates that students do homework near end of the week. On the other hand, lecture videos, lecture questions, and the textbook are not used more frequently over the weekends.

Several trends emerge over the length of the course. Lecture question events decay early as homework activity increases. Textbook use peaks during exams, and there is a noticeable drop in textbook activity after the midterm [18], offset by a rise in lecture videos. The use of discussion forums to help with homework lags the increase of homework activity early in the course.

4.2 Time on Tasks

Time represents a reasonable interpretation of a cost function for students, making it important to study how they allocate time among available course components [15, 19]. Figure 3 shows that the most time is spent on lecture videos. The biggest change over the first seven weeks is the apparent transfer of time from the lecture questions to the homework (see Fig. 3). Spending so much time on lecture questions (that didn’t count toward the grade) is hard to rationalize from a performance goal orientation (5); however on mastery-oriented grounds, it is possible that students saw completion of the longer homework problem assignments as sufficient evidence of understanding the lecture content. However, it is possible that students decided in advance how much time to spend, and had time for lecture questions only when the homework load was light.

The prominence of time spent in the discussion forums is noteworthy as these were neither part of the course sequence, nor did they count for credit. Presumably students spent time in the discussion forums because of their utility in helping with the homework, and also possibly because of their social value. The small spike in textbook time at the midterm, a larger peak in the number of accesses (see Fig. 3), and the decrease in

textbook use after the midterm are typical of textbook use when online resources are blended with traditional on-campus courses [18]. Further studies comparing blended and online textbook use may also prove fruitful [3,17].

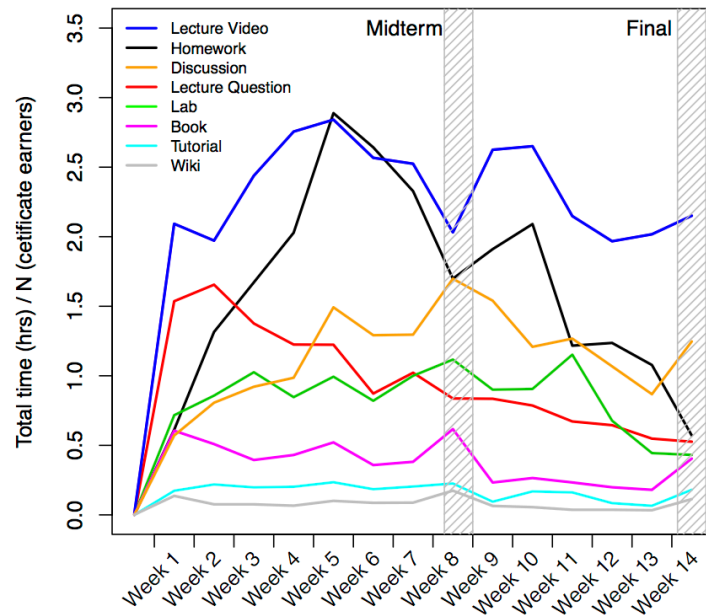


Fig. 3: Time on tasks.

Certificate earners average time spent in hours per week on each course component. Midterm and final exam weeks are shaded.

4.3 Percentage Use of Course Components

The extent to which students use various course components is, along with student time allocation, an important metric for instructors deciding how to improve their courses and for researchers studying the influence of course structure on student activity and learning. For each course component, Fig. 4 shows the percentage of certificate earners that accessed at least the percentage of that resource shown on the x-axis. Homework and labs (each 15% of overall grade) display high fractional usage. These curves drop sharply around 80%, probably reflective of the course policy of dropping the two assignments with the lowest grades. The low proportionate usage of textbook and tutorials is similar to the distribution observed for supplementary (that is, not explicitly included in the course sequence) e-texts in large introductory blended physics courses [16], though the 6.002x textbook was assigned in the course syllabus. The course authors were disappointed with the low usage of tutorial videos and suspected that placing tutorials after the homework and laboratory (for which they were meant to help) in the course sequence was partly responsible. (The wiki and discussion forums had no defined number of resources and so are excluded here.)

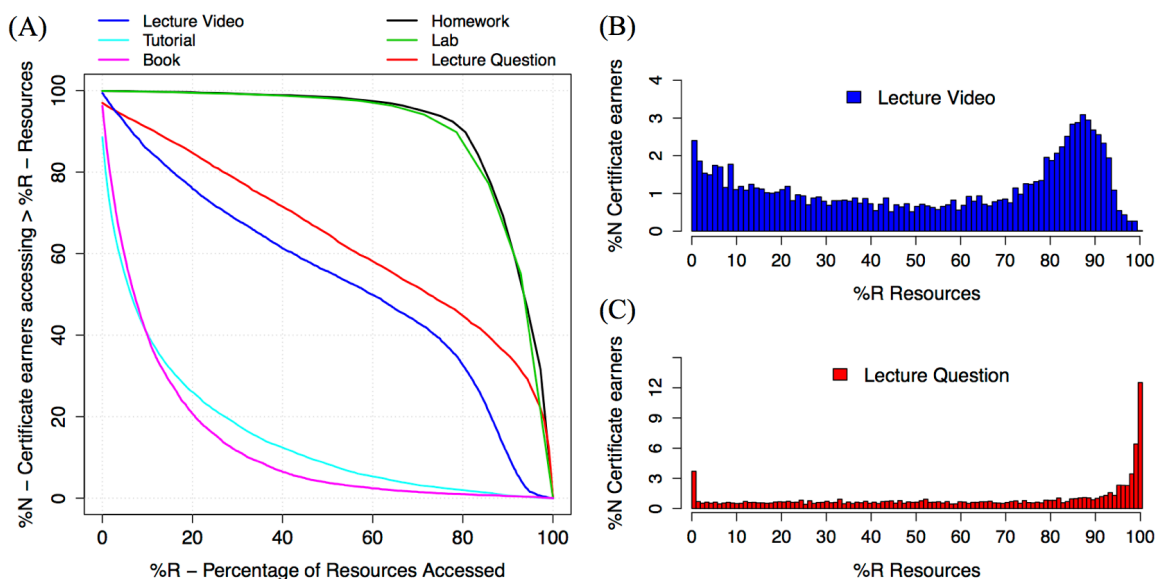


Fig. 5: Fractional Usage of Resources. (A) The percentage of certificate earners that accessed greater than %R of that type of course resource. The density of users is the negative slope of the usage curve. (B) The bimodal distribution for videos accessed (as percentage) (C) Distribution for the lecture questions.

To better understand the middle curves, representing lecture videos and lecture problems, it helps to recall that the negative slope of the fractional resource usage curve is the density of students accessing that fraction of that course component (see Fig. 3b and 3c). Interestingly, the distribution for the lecture videos is distinctly bimodal: one quarter of the certificate earners accessed less than 20%, and 40% accessed over 75%. This bimodality merits further study into learning preferences: are some students learning from other resources exclusively, or are participants who already have good knowledge of the content prior to this course simply not watching the lecture videos? The distribution of lecture problem use is flat between 0% and 80%, and then rises sharply, indicating that 40% of the students access over 80% of them. Although the time spent on lecture problems dropped significantly the first third of the term (See Fig. 3), 75% of the students accessed over 1/3 of the lecture problems, suggesting a trend of looking quickly at these questions rather than ignoring them.

The results shown so far have many hallmarks of effects seen in on-campus studies of how course structure affects resource use [18], and performance outcomes [4,11] in introductory (college) courses. Future MOOCs might extend the large amount of research on education generally, and take advantage of some of its insights (e.g. that frequent exams drive resource usage and maximize learning outcomes [11]).

5. Conclusions

This paper illustrates several methods for analyzing and interpreting the logged data for students who obtain certificates in a MOOC: distribution of their time amongst resources,

their activity overall and for each resource, and their fractional use of those resources. Certificate earners invested the plurality of their time in lecture videos, although one quarter of them watched less than 20%. This suggests the need for a follow-up investigation of the correlations between resource use and learning. Finally, we highlight the significant popularity of the discussion forums in spite of their being neither required nor included in the navigation sequence. If this social learning component played a significant role in the success of 6.002x, a totally asynchronous alternative may be less appealing, at least for a complex topic like circuits and electronics.

In concluding, we re-emphasize our main point: MOOCs are ideal vehicles for education research; partly because they lack the constraints of culture, economics, and classroom design present in on-land education, but mostly because they allow research based on detailed insight into all learning activities of the students. To the extent that the advantage of time-stamped logs of essentially all student behavior and the associated learning throughout the entirety of a course is dissipated by the increased variability of students in MOOCs, the large sample sizes will enable the study of specific student cohorts (e.g. based on demographics [9], effort, and learning habits.). Combining time-on-task observations with measures of learning will allow measurements of the amount learned per unit time spent on a given course component; possibly extending previous studies of online learning [7,15]. We are optimistic that the data from MOOCs can lead to a process of cyclic improvement based on research development, experimentation, and measurement of learning outcomes, allowing substantial improvement of educational content and delivery.

6. Acknowledgements

This work was supported, but is not endorsed, by NSF grant DUE-1044294. We thank MITx for data access, and J. deBoer and other members of the Teaching and Learning Laboratory and RELATE groups at MIT for helpful suggestions and comments. This paper is a shorter and simplified version of a longer paper submitted to Communications of the American Computing Machinery.

7. References

1. Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of educational psychology*, 80(3), 260.
2. Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
3. Cummings, K., French, T., & Cooney, P. J. (2002). Student textbook use in introductory physics. In *Physics Education Research Conference*.
4. Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in Introductory Biology. *CBE-Life Sciences Education*, 10(2), 175-186.
5. Guzdial, M. (1993). Deriving software usage patterns from log files. *Tech Report GIT-GVU-93-41*.
6. Hyman, P. (2012). In the year of disruptive education. *Communications of the ACM*, 55(12), 20.
7. Jiang, L., Elen, J., & Clarebout, G. (2009). The relationships between learner variables, tool-usage behaviour and performance. *Computers in Human Behavior*, 25(2), 501-509.

8. Johnstone, S. M. (2005). Open educational resources serve the world. *Educause Quarterly*, 28(3), 15.
9. Kolowich, S. (2012). Who Takes MOOCs?. *Inside Higher Ed*.
10. Kortemeyer, G. (2009). Gender differences in the use of an online homework system in an introductory physics course. *Physical Review Special Topics-Physics Education Research*, 5(1), 010107.
11. Laverty, J. T., Bauer, W., Kortemeyer, G., & Westfall, G. (2012). Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!. *Physics Teacher*, 50(9), 540.
12. Martin, F. G. (2012). Will massive open online courses change how we teach? *Communications of the ACM*, 55(8), 26.
13. McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). The MOOC model for digital practice. *SSHRC Knowledge Synthesis Grant on the Digital Economy*.
14. Minaei-Bidgoli, B., Kortemeyer, G., & Punch, W. F. (2004). Enhancing Online Learning Performance: An Application of Data Mining Methods1. *Immunohematology*, 62(150), 20-0.
15. Morote, E. S., & Pritchard, D. E. (2009). What course elements correlate with improvement on tests in introductory Newtonian mechanics?. *American Journal of Physics*, 77, 746.
16. Pappano, L. (2012). The Year of the MOOC. *The New York Times*.
17. Podolefsky, N., & Finkelstein, N. (2006). The perceived value of college physics textbooks: Students and instructors may not see eye to eye. *The Physics Teacher*, 44, 338.
18. Seaton, D. T., Bergner, Y., Kortemeyer, G., Rayyan, S., Pritchard, D.E. (*In preparation for submission*). **When and how much do students read e-texts in introductory physics courses?**.
19. Stewart, J., Stewart, G., & Taylor, J. (2012). Using time-on-task measurements to understand student performance in a physics class: A four-year study. *Physical Review Special Topics-Physics Education Research*, 8(1), 010114.
20. Vardi, M. Y. (2012). Will MOOCs destroy academia?. *Communications of the ACM*, 55(11).
21. Warnakulasooriya, R., Palazzo, D. J., & Pritchard, D. E. (2007). Time to completion of web-based physics problems with tutoring. *Journal of the experimental analysis of behavior*, 88(1), 103.