

# Towards Real-Time Analytics in MOOCs

Daniel T. Seaton<sup>1,2,\*</sup>, Yoav Bergner<sup>2</sup>, Isaac Chuang<sup>2,3</sup>, Pitor Mitros<sup>4</sup>, and David E. Pritchard<sup>2</sup>

<sup>1</sup> Office of Digital Learning, <sup>2</sup> Department of Physics and RLE, <sup>3</sup> Department of Electrical Engineering and Computer Science, <sup>4</sup> edX and CSAIL,  
Massachusetts Institute of Technology, Cambridge, MA 02139

(\*Correspondence to: [dseaton@mit.edu](mailto:dseaton@mit.edu))

**Abstract.** Massive open online courses (MOOCs) collect essentially complete records of all student interactions in a self-contained learning environment, with the benefit of large sample sizes. Building on our data mining of the first course in MITx (now edX) we demonstrate ways to analyze data to illustrate important issues in the course: how to distinguish browsers from certificate-earners, which resources were accessed the most and how much time was allocated by certificate-earners. Each topic is addressed via appropriate displays that, in future courses, can be updated in real time. Furthermore, we stress that analytics can provide useful information to teachers, to resource creators (authors), and to members of organizations trying to improve their MOOCs.

## 1 Participation: Attrition, Tranches, and Total time.

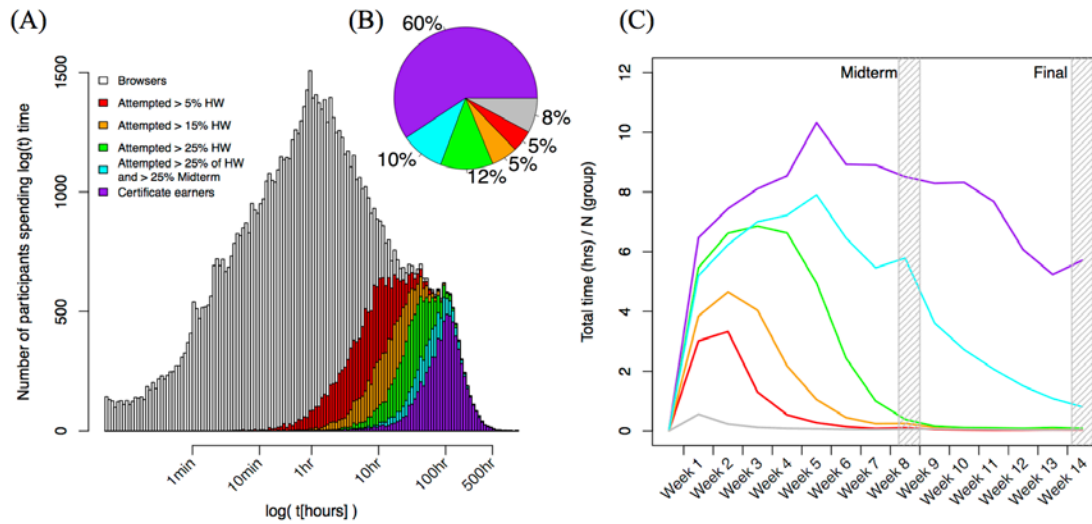
Participation in the inaugural edX course, 6.002x: Circuits and Electronics [1], had over 100,000 registrants. Not only was this population large, but it consisted of people from a myriad of cultural and educational backgrounds. Our goal has been to extract as much meaningful information as possible from tracking logs [2] and limited user profile information in 6.002x, highlighting both research results [3] and analytics that could potentially inform future MOOCs and traditional on-campus instruction.

Without access to participant's demographic profiles, we chose to categorize participants based on their level of interaction with assessment items. In Fig. 1, we show how much time participants spent in the course highlighting 6 tranches (slices): Browsers who did not attempt any assessment, 4 tranches based on amount of assessment attempted, and of course, certificate earners. In order to better understand our tranches, we plot the total-time spent per week by each tranche. These results show that time per week has a significant relationship with attrition.

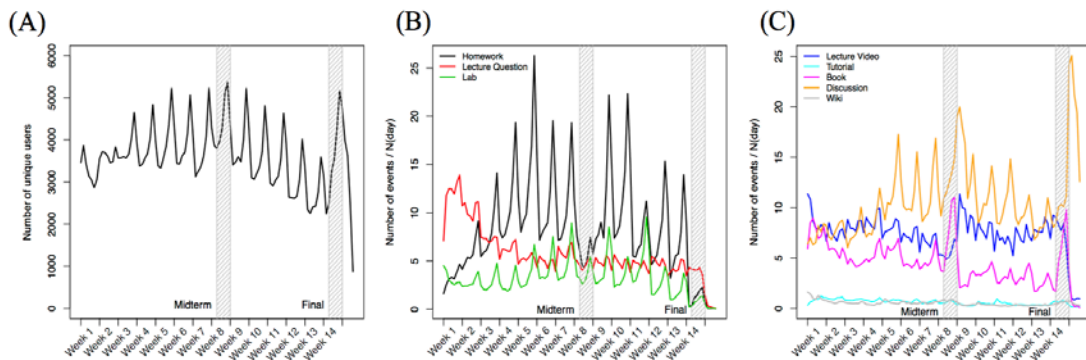
## 2 Frequency of Accesses Illustrates Activity.

The number of accesses of various course components in 6.002x by those active each day, which we call activity, is shown in Fig. 2. Homework and laboratory activity has consistent periodicity with the number of unique users per day, while lecture questions do not share this periodicity and also suffer a downward trend in

overall activity for the term. For learning based components, Discussion activity is the only component sharing periodicity with the unique users per day, suggesting a correlation with for-credit activity. These observations may imply that students working on graded activities do not utilize many learning-based resources.



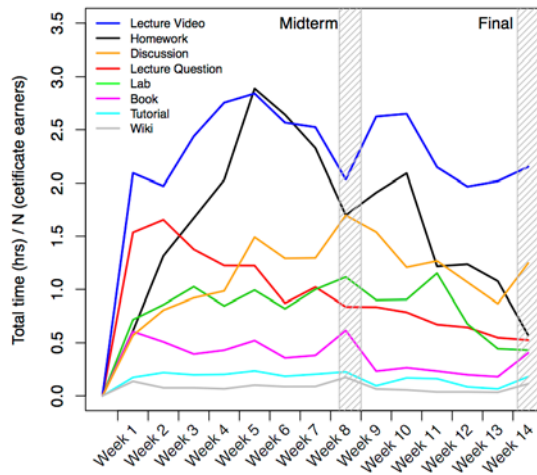
**Figure 1:** (A) Distribution of time spent by 108k participants in the course (~7200 received certificates). We have divided the non-certificate earners into tranches (colors) based on the percentage of assessment items they attempted. (B) Percentage of total measured time in the course, and (C) average time per student per week for each tranche.



**Figure 2:** From left to right, number of unique certificate earners 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 weekly periodicity. Assessment-based components activity per active student each day (B) shows end-of week periodicity of for-credit assessment (homework and labs). Learning-based components (C) show that discussion forums alone display strong periodicity and that textbook activity drops after the midterm while lecture video activity rises. Textbook activity rises dramatically during exams.

### 3 Time on Various Resources.

Time represents the principal cost function for MOOC students, and it is therefore important to study how students allocate time among available course resources. Of the course components offered in 6.002x, Lecture Videos and Homework generally took the most time each week. Discussion Boards (which were voluntary) represent the next highest level of time allocation by students. It is also interesting that the Discussion time trends upward relative to homework time for later weeks in the course, suggesting increased use by students doing homework. A downward trend is observed for Lecture Questions. Other course components have consistent time across the course, but appear to have minimal activity.



**Figure 3:** Average total-time spent by certificate-earners on course components in 6.002x. Time is aggregated over the week in which each module was due.

### 4 References.

1. 6.002x: Circuits and Electronics. - <https://6002x.mitx.mit.edu/>
2. Guzdial, M. (1993). Deriving software usage patterns from log files. *Tech Report GIT-GVU-93-41*.
3. Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., Pritchard, D.E. (2013). Who Does What in a Massive Open Online Course? *In Press*. - *Communications of the ACM*.