**Wellesley College**
# Wellesley College Digital Scholarship and Archive

Faculty Research and Scholarship

2014

# What does enrollment in a MOOC mean?

Eni Mustafaraj

Follow this and additional works at: http://repository.wellesley.edu/scholarship

**Version: Publisher's version**

## ABSTRACT

In 2012, when MOOCs became largely known, media reports were fascinated with the big number of enrollments. The number 150,000 students was mentioned for both Stanford's *Artificial Intelligence* course and MIT's *Circuits and Electronics*, to be later followed by the underwhelming completion rates, that often are in the single digit percentages[1]. But what kind of enrollment do these large numbers really show? We try to answer this question by breaking this number into its components, while comparing two successive iterations of the same MOOC offered on the edX platform.

### Author Keywords

MOOCs; learning analytics; engagement; visualization

### ACM Classification Keywords

H.5.3 Evaluation/methodology

## INTRODUCTION

For this study, the author was given access to anonymized data from two iterations of MITx 6.00x – *Introduction to Computer Science and Programming*. Each course ran over a 4-month period: Fall 2012 (Oct 2012 - Jan 2013); Spring 2013 (Feb - May 2013). The stated prerequisite was "high school algebra and a reasonable aptitude for mathematics", and the estimated effort per week was around 10-12 hours. Both courses were hosted on the edX platform. The Fall 2012 offering was one of the first official edX courses, since the first MITx course (6.002x *Circuits and Electronics*[2]), was offered during a period in which edX wasn't established. As such, 6.00X (Fall 2012) attracted a considerable amount of curiosity and attention, reflected in the large number of sign-ups for the course, which we have estimated to more than 184,000 students. But, as we will demonstrate in this paper, this number doesn't reflect the behavior of the participants in the course, and therefore shouldn't be taken into consideration for further purposes of evaluating course success.

---

[1] **http://www.katyjordan.com/MOOCproject.html**
[2] **https://6002x.mitx.mit.edu/**

## DATA DESCRIPTION

Data for each course in the edX platform are stored in different databases: course content, discussion forums, student personal data, student course progress, event tracking, etc. The analysis in this paper is based only upon data from the event tracking database. A database entry uses the JSON format[3], as shown in the (truncated) example below:

```
{ "username": "123456",
"event_source": "server",
"event_type": "/courses/MITx/6.00x/2012_Fall/info",
"time": {"$date": 1348897556438},
"event": {"POST": {}, "GET": {}},
... }
```

Although this data allows us to recreate the complete history of a user interaction with the course content, in this paper we will only focus on **when** and **how often** students showed up for the course. The nature of their activity on the website is the topic of another paper.

## DATA ANALYSIS

By parsing the dates and event types of every user in the database, we are able to find out when, how often, and what they did in every website visit. The largest group of users (84,853 users or 46% ) had a single visit to the website in Fall 2012, which corresponded to landing on the info page after clicking the "Sign-up" button. As observed elsewhere [1], one can better compare such an action to a Facebook Like event, than to the real intention to enroll in the course. Following a classification proposed in [2], we will label these users, whose only activity was signing up, as "no-shows". No-shows continue their sign-up routine during the entire course duration, as well as after the course is closed[4].

Another large group is that of "one-day visitors". This group consisted of 19,035 (Fall 2012) and 21,615 (Spring 2013) such users, who spent a median time of eight consecutive minutes on the site. A visualization of the number of unique daily visits by all users can be found online (see footnote 4). The majority—78% of all users (excluding "no-shows")—visited between 1 to 10 days during the duration of 112 days.

## DISCUSSION

These results allow for a series of observations:

---

[3] **http://en.wikipedia.org/wiki/JSON**
[4] Due to limited space, most of visualizations for the data analysis in this paper can be found online at: **http://cs.wellesley.edu/~eni/mitx/**.
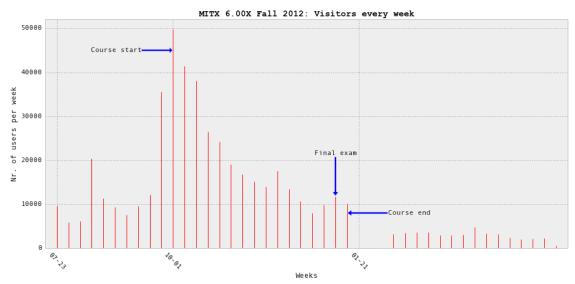
Figure 1: Weekly visits by users during the entire run of the Fall 2012 course. The website was opened for sign-ups 10 weeks in advance of the start date. The first week of course (Oct 1st, 2012) had the largest number of visitors: 49,841. In the week of the final exam, the website was visited by 11,767 users. The website was blocked for a few weeks, and then reopened again.

*Sign-ups might indicate interest in the topic, but not necessarily intention to attend the course.* In fact, while the Fall 2012 offering attracted 184K users, only about 50K showed up in the first week of the course (see Figure 1). Meanwhile, the Spring 2013 offering (only three weeks after the completion of Fall 2012), attracted 90K users[5]. This suggests that successive offerings of the same course might be "less" massive.

*Sequential Enrollments.* The two sets of students enrolled in the two course offerings are not mutually exclusive. In fact, 33,351 users signed-up for both. Informal discussions among participants in the website Reddit[6] indicate that some users were signing up again to provide help in the discussion forum, while others to repeat and reinforce the material. We will look at the behavior of such group in our future research.

*Opening enrollments when there is no content might be damaging.* The large number of sign-ups that became 'no-shows', especially in the Fall 2012 offering happened in the ten weeks preceding the course start, when there was no course content on the website.

*Completion rate or engagement rate?* Media and critics have been measuring completion rates of MOOCs as the percentage of the signed-up students who received a completion certificate. For a course such as 6.00X Fall 2012, with 184,234 signups, where 46% of users never showed-up, the number of students who attempted the final exam (7,559) is really small, only 4%. However, if we base calculations on who showed up for the course on at least 1/4 of its duration (on four different weeks during 16 weeks; there are 35,173 such users, refer to online plots), the completion rate increases to 21%. Fi-

nally, real engagement might be found in the group of 7,161 students, who showed up for more than 40 days during the entire course duration, and all attempted the final exam.

The discussion in this paper is based on **when** and **how often** students visited the course website. Other researchers have advanced a terminology for categorizing users based on engagement with the course material, such as: browsers [4], samplers [3], or observers [2]. We will look into this topic in future research.

### REFERENCES
1. Bruff, D. Lessons Learned from Vanderbilts First MOOCs, 2013. `http://cft.vanderbilt.edu/2013/08/lessons-learned-from-vanderbilts-first-moocs/`.

2. Hill, P. Emerging Student Patterns in MOOCs: A (Revised) Graphical View, 2013. `http://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view`.

3. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proc. of LAK '13*, ACM (2013), 170–179.

4. Seaton, D., Bergner, Y., Chuang, I., Mitros, P., and Pritchard, D. Who does what in a massive open online course? *Communications of ACM* (Forthcoming).

---

[5]This number is not exact, since there are several missing days in the database.

[6]`http://www.reddit.com/r/600x/`