Understanding the edX MOOC: How can “Circuits and Electronics” (6.002x) help us understand the MOOC learning experience?

Dr. Lori Breslow, MIT Teaching and Learning Laboratory, Professor David Pritchard, MIT Department of Physics and RELATE; and Professor Andrew Ho, HGSE

Project Summary

Open-access online courses have recently expanded very publicly, as high-profile research universities in the U.S. throw open their virtual doors to students around the world by developing massive open online courses (MOOCs). While this is being done with much fanfare, the educational implications have not been thoroughly investigated. The characteristics of the students, their interactions online (both with the course and each other), and the factors that contribute to their academic success in MOOCs have only been broadly ascertained.

This one-year RAPID grant will allow educational researchers from MIT and Harvard to investigate “Circuits and Electronics” (6.002x), the first offering of edX, the MIT/Harvard collaboration for online learning. This 14-week course, which began in spring 2012, delivered content via two interactive video sequences per week interspersed with online exercises. These were accompanied by weekly textbook readings, homework, and labs. The course site contained a Wiki where students could share new knowledge as they progressed through the course, as well as a discussion board where students could have their questions answered by fellow students or paid teaching assistants.

In all, 154,763 students registered for 6.002x over the life of the course. Data about these users include the locations where they accessed the course via the IP address, as well as behavioral data any time they interacted with the course. For example, data from access to the online textbook and videotaped presentations show when and how long access windows were open. Data from the interactive exercises tell us which answer(s) were chosen by the student, number of attempts, and number correct (some questions were open response and some were multiple choice). Additionally, we have student postings from the discussion boards. Grades were determined by scores on three problem sets, a midterm, and a final exam, which are standard for “traditional” MIT courses. We have performance data from these assessments—number of attempts on various problems, how long the assessments were open, and overall scores. Finally, MIT Teaching and Learning Laboratory researchers developed and MITx developers distributed a survey to students who completed the course to get a more detailed picture of this sub-sample of the student population.

We will use these data to answer the broad questions, “Who are the students who enrolled in 6.002x, how did they use it, how do these factors relate to their learning outcomes, and how did 6.002x interact with the residential experience?” We will begin this work by examining student background (e.g., country of origin) to develop a profile for the overall population of students who accessed the 6.002x site; we will then analyze, in more detail, the demographics of the students who completed the course. Next, we will investigate learning behaviors (e.g., time on task for different course components) in 6.002x. Our goal is to see if there are

* Contact information: Dr. Lori Breslow, Director, MIT Teaching and Learning Laboratory, lrb@mit.edu. This project is funded by the National Science Foundation, Division of Education and Human Resources, grant number DRL-1258448.
relationships between student background and/or learning behaviors and persistence (defined as completion of the course) and achievement (defined as the total points earned on the final exam). We will then explore the learning communities and the interactions within them, and, again, their relationship to persistence and achievement. Lastly, we will study the interaction between online learning and residential learning in the same course. The objective of this research is to identify the types of instructional materials and strategies that optimize learning outcomes for groups of learners who may differ in age, level of preparedness, family or work responsibilities, etc.

**Intellectual merit.** The intellectual merit of the project is, first, in its employment of a novel dataset that can inform teaching and learning in the rapidly growing, more potentially equitable online classroom. Because 6.002x gathered data on learner activities at the level of seconds, the project offers a novel scope for the use of survival analysis in higher education. The data also allow us to analyze the relationship between student background and both individual and group behaviors in a more rigorous way than has previously been possible. Finally, our work will contribute to the body of knowledge about online learning by estimating the predictive power of student characteristics and behaviors on course persistence and academic achievement.

**Broader impact.** We expect the findings from this study will be generalizable to other populations of students who engage in online learning. As edX expands its offerings, our research will allow content experts, instructional designers, and developers to improve subsequent courses and instructional practices. Initiatives like edX will allow access to courses and communities offered by highly ranked institutions to the groups most likely to face barriers to entry, including underrepresented populations. Understanding how students in different formal and informal spaces utilize edX courses will allow educators to use its materials more effectively on the residential campus as well.
Project Description

I. Significance

Despite major international efforts such as the Millennium Development Goals and Education For All (UNESCO, 2012), educational opportunity remains highly stratified around the world. While the advent of the Internet may have greatly increased access to information (International Telecommunications Union [ITU], 2011), and while MIT has published course materials online for over 2000 courses (OCW, 2012), access to the teaching and learning in an elite university has continued to elude the vast majority of residents of the supposedly flat world.

Since the late 1990s, the two-way communication capability of the Internet has allowed for more formal exchanges between teachers and their students. Online courses are not new, but the affordances and constraints they provide for the learner are not well understood. Major research institutions with highly selective traditional residential programs are branching out into this new space, with investment proliferating in massive open online courses (MOOCs). MIT and Harvard partnered to create edX, an online learning initiative, in May 2012, and the University of California-Berkeley joined the collaboration in July 2012.

MITx launched its first MOOC, 6.002x, in the spring of 2012. This 14-week course delivered content via two interactive video sequences per week accompanied by weekly textbook readings, homework, and labs. The interactive video sequences included online exercises that students could complete to check their understanding, and there were optional tutorials that demonstrated how to solve 6.002x problems or complete labs. The course site contained a Wiki, where students could share new knowledge as they progressed through the course, as well as a discussion board where students could have their questions answered by fellow students or paid teaching assistants.

Although research to date has provided educators with some understanding of the characteristics of online learners, best practices for delivering online instruction, and barriers to course completion, the data amassed from a widely accessible course such as 6.002x have the potential to answer major questions surrounding open-access online learning. As well, the range of technical options for presentation of content is expanding rapidly and may result in significant alterations to our present model of online education. Additionally, the quality of open access online courses offered by highly selective institutions such as MIT, Harvard, and Berkeley may change learner demographics. These factors have the potential to challenge our current knowledge about online learning and instruction in ways that need to be explored.

II. Objectives and Research Questions

In this proposal, we describe a highly feasible, high-priority study that will provide the first steps in answering a broader set of research questions. Specifically, this study will examine the particular nature of MOOCs and the students who interact with and learn from them. It will focus broadly on the questions, “Who are the students who enrolled in 6.002x, how did they use it, how do these factors relate to their learning outcomes, and how did 6.002x interact with the residential experience?”

We will begin this work by examining student background (e.g., country of origin) and learning behaviors (e.g., time on task for different course components) in 6.002x to see if there are relationships between these variables and persistence (defined as completion of the course) and achievement (defined as the total number of points earned on the final exam). We will then explore the learning communities and the interactions within them as well as their relationship to persistence and achievement. Lastly, we will explore the interaction between online learning and residential learning in the same course. The objective is to identify the types of instructional materials and strategies that optimize learning outcomes for groups of learners who may differ in age, level of preparedness, family or work responsibilities, etc.
The following are the specific questions we seek to answer. For each question, we describe what we will study, as well as how the results can inform future courses and a research agenda for MOOCs.

**RQ1: What characteristics of users explain the variability in achievement?** What are the characteristics of different users in the course? How do their background characteristics and interactions with course components predict achievement?

Educational research has identified the predictors of academic success in traditional learning environments, including: individual characteristics such as age, gender, or race/ethnicity; home factors such as parental education or income; community factors such as urbanicity. However, the predictive power of these factors has not been clearly ascertained in university-level MOOCs. We will first identify the descriptive characteristics of the sample participating in 6.002x, and then we will estimate the explanatory power of these factors for achievement.

Another stream of research in post-secondary education has identified pedagogical practices that foster mastery of knowledge and skills, such as interactive engagement (e.g., Smith et al., 2005), and that promote retention and transfer, such as practice at retrieval (Halpern & Moskel, 2003). Technology offers the promise of strengthening these successful pedagogical techniques, but, as yet, we have not determined the most effective ways to do that. This study will allow us to identify the use of course components that may be beneficial for student learning.

The components of 6.002x that have been shown to be related to high achievement can be further investigated to support the development of edX and other MOOCs. For example, we can explore interactive engagement by estimating the impact of time spent working on online labs. We can study the effect of practice at retrieval by looking at the relationship between scores on practice problems and the final exam score in the course. Components that are shown to be useful for certain types of learners can be optimized in future courses, and those negatively associated with higher achievement can be improved. Components whose estimates suggest they are related to higher achievement can then be tested as isolated interventions in randomized control trials to determine possible causal effects.

**RQ2: What characteristics of learners and the ways in which they interacted with course components predict their completion of the course?** What characteristics of users explain the variability in a hazard model of their persistence throughout the 14 weeks? What level of interaction with course components predicts persistence throughout the 14 weeks?

We will estimate the predictive power of student background characteristics and interaction with course components for persistence in 6.002x. Most investigations of student dropout in traditional learning environments look at a student’s trajectory over the course of a degree program or an entire academic career. Because data were collected more frequently in 6.002x, we can track users as they progressed through the course, and we can see when they chose to stop their participation. There were a significant number of dropouts in 6.002x, with less than 5% of the students who signed up at any point completing the course. While we realize there were many reasons people enrolled (including being simply curious), we seek to identify the factors that might predict dropout for serious students. We believe this will help identify at-risk users who need additional tailored support in the online environment. Future edX courses can then provide those services while at-risk users are still enrolled.

**RQ3: What is the composition of online learning groups and how do they support course persistence and achievement?** What are the characteristics of groups that emerged organically in 6.002x? How did the groups, as a unit, interact with course components? How does group interaction relate to individual persistence and achievement? What course components were used by learning groups in which individual members persisted and achieved academic success?
The interactive nature of the 6.002x portal allowed students to find and make use of virtual communities. Previously, one of the main arguments against online courses, in comparison to traditional classrooms, was the lack of peer/community interaction. The architecture of MITx, however, allows unprompted student discussion. But researchers do not know what kinds of communities this new interactive medium facilitates. We are interested in how students identify with or create organic learning communities in a large space like 6.002x.

We will first identify the groups that formed during 6.002x, and we will investigate their distinguishing characteristics (e.g., frequency of communication). We will then look for relationships between those characteristics and the persistence and achievement of student members. Finally, we will examine individual levels of interactivity within the group to see if they are predictive of that individual’s persistence and level of achievement. Once the benefits of particular types of groups and levels of interaction are identified, we can guide future students to groups that will aid their achievement in courses like 6.002x.

**RQ4: How does the online student experience interact with the residential learning experience?** How do residential students describe their use of 6.002x? How do they interact with the course material? Does academic achievement differ between 6.002x students who did or did not have additional in-person experiences? Does academic achievement differ between 6.002x students who had additional in-person experiences and students who took the on-campus version of the course (which is numbered 6.002)?

Of substantial interest to colleges and universities around the world is how MOOCs will influence the nature of their residential programs. A small subgroup of 6.002x students was also enrolled at MIT as residential students. This group was given the opportunity to augment their online experience with face-to-face interactions such as labs and recitations. From an examination of the learning behaviors of these students, along with a comparison of their performance with the performance of students enrolled in the traditional residential version of the course (i.e., 6.002), we can gain insight into the benefits that in-person experiences provide to online learners, as well as the benefits that online experiences might provide to residential students.

Findings from the above four research questions will provide data to be used to modify the edX platform and MOOCs, more generally. Student characteristics or behaviors strongly associated with success will provide input to improvements in curricular materials and pedagogical strategies.

**III. Current Knowledge about Online Learning**

Since its inception, online learning has been studied to understand the characteristics of the students who utilize it, as well as its relationship to student learning outcomes. We summarize some of the work that is relevant to our proposed study below and explain what our work will add to this body of knowledge.

**The Online Learner**

Online learners are currently characterized as being primarily female and adults between the ages of 25 and 50. These are students who want a flexible, cost-effective way to gain or update skills (Park & Choi, 2009); who are seeking an affordable means to attend college outside of their home country (Crawley, 2012); or who are younger post-secondary students who grew up in relative economic security and view technology as routine in their everyday lives (Dziuban et al., 2007). Disparities in Internet access (ITU, 2011) and use are apparent internationally as well as locally (DeBoer, 2012). We argue that the widespread availability of MOOCs will change access to educational materials, which, in turn, will significantly change the demographic characteristics of online learners around the world. Our work will undertake the first important step in identifying the demographic characteristics of this new group of learners in order to better support their educational needs.
**Effectiveness of Online Learning**

A recent meta-analysis showed that, on average, students learning online perform better than students learning the same material through traditional or face-to-face instruction (U.S. Department of Education [USDE], 2010). Results from this work also suggest that the educational features available in an online environment are not significantly responsible for improved student learning outcomes. Instead, learning is improved when students are offered opportunities for collaboration, given more control over their engagement with online materials, and provided mechanisms to reflect on their understanding of the material (USDE, 2010). Our use of a more fine-grained approach to identify students’ online behaviors that predict their academic success will add to the current understanding of factors that optimize online learning.

**Attrition of Online Learners**

The increasing number of students who enroll in online courses does not positively correlate with the large number of students who successfully complete those courses, however. In trying to understand this problem, researchers have identified the following barriers: difficulty in accessing electronic libraries or technical support; feelings of isolation from faculty and peers (Bunn, 2004); poor communication with faculty (Aragon & Johnson, 2008; Bunn, 2004); decreased engagement in course activities (Morris et al., 2005); and non-academic issues (e.g., illness [Aragon & Johnson, 2008; Bunn, 2004; Park & Choi, 2009]). Our work will provide evidence from the examination of learning behaviors to expand our understanding of the causes of attrition in online learners.

**Online Learning Communities**

Participation in interactive learning communities may address some of the possible causes of student attrition in online courses (Angelino, Williams, & Natvig, 2007; Hart, 2012). However, not all students make use of them. Features such as the course and user interface, as well as previous experience with online learning communities, shape student perceptions of and intentions to use communities as a resource (e.g., see Liu et al., 2010). The potential for MOOCs to attract an extremely diverse user population makes language, culture, and previous online experience important considerations for designers as they contemplate how to entice students to join online communities (Thorne, Black, & Sykes, 2009). Our work will further explore the significance of characteristics of online learning communities in predicting academic outcomes.

**IV. Methods**

**Data**

6.002x data contains observations from a variety of points in time. At our first time point, when the course went live, 120,000 users registered. However, over the entire life of the course, 154,763 users registered. Of those, 69,221 accessed the first problem set, and 26,349 earned a score of one or greater on that problem set. The midterm was accessed by 13,569 users; 10,547 earned a score of one or greater and 9,318 passed. The final exam was accessed by 10,262 users; 8,240 earned a score of one or greater and 5,800 users earned a passing score. The course was passed by 7,157 students.

6.002x user data includes the locations where students accessed the course via the IP address, as well as behavioral data any time the student interacted with the course. For example, data from student access to the online textbook and videotaped presentations show when and how long access windows were open. Data from the interactive exercises that accompanied the video sequences tell us which answer(s) were chosen by the student, number of attempts, and number correct (some questions were open response and some were multiple choice). Additionally, we have student postings from the discussion boards. Grades were determined by scores on three problem sets, a midterm, and a final exam, which are standard for “traditional” MIT courses. We have performance data from these assessments—number of attempts on various problems, how long the assessments were open, and overall scores. Finally, Teaching and Learning Laboratory researchers developed and MITx developers distributed a survey (using matrix sampling) to course completers to get a more detailed picture of this sub-sample of the student population.
**Analysis**

The first step in our analysis will be to generate descriptive statistics for all demographics. Next, we will estimate the means for student behaviors such as time on task in each of the course components. We will then construct a multi-level model (MLM), nesting students within sub-national units (identified via IP address) within countries.

In order to explore achievement, we will use the multi-level model to investigate the variance in achievement (cumulative final exam score) that can be attributed to country and city, or other municipal units within country. We will also explore student behaviors, such as time on task for various course components, to determine if there is any relationship between those behaviors and achievement. As well, we will compare the predictive power of component scores—which have no weight in the final exam score themselves—for predicting the final outcome score.

In order to explore persistence, for each time students were assessed (i.e., the three problem sets, midterm, and final), we will determine which students dropped the course and which ones continued. We will then create a hazard model using student-level demographics and longitudinal behaviors to predict dropout from the course. Survival analysis will help us not only to identify the factors associated with attrition, but also to isolate factors that predict attrition at particular times; this is important for MOOCs, where attrition immediately after signup is notably different from attrition at the end of the course. We will make a methodological contribution to the educational literature by estimating a hazard function that investigates student dropout rates at specific points within the course, a much finer-grained analysis than the hazard models that are currently reported.

For students who completed the course, we will include additional demographics, and we will determine whether the estimates of the variance in persistence and achievement within and between countries and the variance explained by student behaviors remain after including factors such as students’ previous degrees and parents’ educational level (these data are available from the TLL survey).

In studying the user groups, we will employ social network analysis to first identify the communities that students created and participated in. We will then use student-level demographics and behavioral information to estimate the predictive power of these characteristics on frequency of interaction and the type of group with which students engaged. Future work can build on these findings about how groups support persistence and achievement.

Finally, we will use both qualitative and quantitative methods to investigate the interaction between online and residential experiences. Interviews will explore the experience of students who took the traditional on-campus course (6.002). We will also create a matched sample of students who took 6.002x with an added on-campus experience at MIT and students who took 6.002x only online. We will use propensity score methods to match 6.002x/residential students with 6.002x-only students who were nearly as likely to have enrolled in the residential experience but did not. We can then compare their final exam scores; assuming that observables used to create the matched sample would address possible sources of bias. We may test the inclusion of controls to verify that the matched sample has less bias.

**V. Intellectual Merit**

This research is the first step in identifying the diverse student population that will be attracted to massive open online courses, as well as best practices for facilitating their learning. A thorough understanding of the relationship between the characteristics of the students, the behaviors they exhibit as they navigate through the course, and their persistence and academic achievement will enable course developers to improve current methods of delivery and support systems. In particular, this work will add to the body of knowledge regarding factors associated with the student attrition from online courses. Our research will also contribute to the understanding of how student-created communities emerge, their use by different
types of learners, and their potential benefits. Lastly, we will begin to build an understanding of how MOOCs will impact the residential experience.

VI. Broader Impacts and Dissemination

We expect the findings from this study will be generalizable to other populations of students who engage in online learning. As edX expands its offerings, our research will allow content experts, instructional designers, and developers to improve subsequent courses and instructional practices. Initiatives like edX will allow access to courses and learning communities offered by highly ranked institutions to the groups most likely to face barriers to entry, including underrepresented populations. Understanding how students in different formal and informal spaces utilize edX courses will allow educators to use its materials more effectively on the residential campus (for example, in the so-called “flipped classroom”) (Young, 2012).

We lay the groundwork for expanded research on the trove of data in the 6.002x database. We plan to expand our research efforts by proposing a larger follow-up study, building on the foundation provided by this grant, to investigate how students engage academically and socially online; leaders in higher education have identified this as an important area of study (e.g., Bacow, Bowen et al., 2012). Our future research will expand our examination of the learning communities associated with 6.002x to understand the quality of student interactions more deeply. We will also continue to investigate the experiences of online learners who do and do not have accompanying in-person experiences. Finally, we will look at the implications for students if a larger, multi-course certificate program were to be developed.

As the first research study done using these data, we will disseminate our results through academic journals, professional publications (e.g., *The Chronicle of Higher Education*), and the popular press. Adhering to the edX approach of providing open source materials, our research will enable future educators to build on the knowledge we have amassed, improving learning opportunities for all students.

References Cited


