

How MOOCs present massive opportunities for research on learning

Gary W. Matkin, University of California, Irvine, USA

ABSTRACT. The *massive* in MOOCs has been limited so far to the number of people enrolled. Little understood or explored is the potential for the analysis of massive amounts of data on learning across thousands of learners. This presentation describes the data being collected by one early university partner in Coursera, how it is being used, and the plans to expand that use beyond what is currently being provided by Coursera. It will also describe the current barriers to more detailed research, what can be done to remove those barriers, and potential of MOOCs to advance learning research beyond existing perceptions of what can be achieved.

KEYWORDS: Coursera, MOOCs, OCW, OER, OpenCourseWare

Introduction

MOOCs have dominated the conversation in higher education since July 2011, when two Stanford professors presented two online, free courses in artificial intelligence. Two years later, nearly four million students have signed up for MOOCs and a number of ventures, both for-profit and non-profit, have been created to offer MOOCs and exploit their potential. There has been a growing and overwhelming public demand for the cost of higher education to be reduced, and many people are looking at MOOCs as a means of achieving this objective. In fact, MOOCs are increasingly being embedded in public policy, state laws, foundation-funded projects, and institutional strategies.

The public discussion about MOOCs has ranged from wild expectations, that they represent an improvement in higher education with the prospect of reaching hundreds of millions of people around the world who are cut off from high quality education, to alarm that MOOCs will disrupt higher education systems, decrease overall educational quality, and even threaten the jobs of the faculty. After two years of this debate, as reality

sinks in, it can be seen that MOOCs are entering a “trough of disillusionment” as several major institutions retreat from the initial enthusiasm with which they greeted MOOCs. Nevertheless, MOOCs and the organizations dedicated to offering them continue to proliferate.

Learning research and innovation dissemination is perhaps the most valuable potential for MOOCs and their offering. The following is a description of how the future of learning research might be influenced by massive open higher education and how learning innovations might be rapidly disseminated and evaluated on a scale never imagined until now. The early experience of the University of California, Irvine (UCI), a Coursera partner and one of the most prominent purveyors of MOOCs, will illustrate both the possibilities and the barriers to the realization of massive learning research and rapid innovation dissemination.

Massive research: possibilities

The Hubble Space Telescope launches scientific breakthroughs

In 1990, NASA launched the Hubble Space Telescope into orbit, making it possible to view the universe unhindered by the atmosphere of Earth. After solving initial problems through a manned repair expedition, Hubble began transmitting data, allowing researchers around the world to use this hugely expensive instrument. Even amateur astronomers were given opportunities to use the telescope through a peer-review process based on scientific merit. Since the launch of Hubble, it is estimated that more than 9 000 peer-reviewed papers and publications have been based on Hubble observations. Hubble-based papers are cited twice as often as other papers. These papers and observations have resulted in significant breakthroughs in astronomy and cosmology, such as confirming that the universe is continuing to expand at an increasing rate, that black holes exist, that many stars have planets, and that those planets have moons of their own. Thus, the launching of Hubble also launched a quantum leap in the science of the universe.¹

1. http://www.nap.edu/openbook.php?record_id=11169&page=16

MOOCs have a similar potential for learning research

One of the systematic barriers to valid and generalizable research in learning is the problem of validation. Since online courses started to become important in higher education around the mid-1990s, a significant amount of energy has been spent on trying to determine the effectiveness of online education in comparison to classroom-based education. There aren't any answers to this question because of the number of uncontrollable variables in the research. The general notion that there is "no significant difference" is now widely held now. But the real conclusion should be that there is not really any valid research available. However, with MOOCs it is now possible to examine the relative effectiveness of alternative learning treatments across large numbers of students. The large sample of students involved can provide a level of statistical validity that is often lacking in smaller-scale studies. And the tremendous diversity of MOOC students can provide insight into learning differences across populations within just one educational treatment. (Of course, an effective way of quantifying the degree of diversity among MOOC course-takers needs to be developed.) "A/B" testing whereby alternative-learning treatments can be delivered to the same population can also be a source of new knowledge about how people learn and what is effective.

The massive nature of MOOCs and the homogeneity of the learning treatment delivered offer the same kind of quantum leap to education researchers that the Hubble telescope provided to cosmologists. Both advances bring many more objects for study into the research arena, both provide researchers from around the world the opportunity to cooperate in their research activities, and both share data that is open and available to all. Let's use a very preliminary actual example of the data that are currently available and how they might be used.

MOOCs' big data and UC Irvine's experience with Coursera

UCI joined Coursera on September 19, 2012, announcing courses on the platform at the same time as 16 other universities, bringing the number of Coursera partners at that time to 33. UCI initially offered six courses, which started in January 2013 and which had, at

one point, more than 250 000 enrollments (at this writing in August 2013, UCI has six additional courses on Coursera scheduled to start in August with current enrollments totaling nearly 50 000). As a matter of routine, Coursera provides UCI with the following data through a query system that allows UCI to access the data at any time:

- List of courses - past, scheduled, and planned (not yet open for enrollment)
- Total enrollment for each course (all enrollment data are updated continuously)
- Number of students in each course who are participating in the Signature Track option.²
- Number of Signature Track students who are receiving Coursera-offered financial aid to cover the Signature Track fee.
- Number of students enrolled in ACE-approved courses.³

In addition, Coursera currently provides large amounts of data that are currently much harder to access, download, manipulate, and interpret. These data include:

- Administrative content for website administration (excludes video content, presentation slides, supplementary educational materials, etc.)
- Instructions for surveys, quizzes, homework assignments, peer-graded assessments, etc.
- Time-stamped copies of student responses for surveys and learning assessments
- Time-stamped logs of student activities (e.g. watching videos, submitting assignments, etc.)
- Discussion forum content including complete text of posts and comments
- Student registration information
- Time-stamped clickstream logs that track student activity on the course website; includes user-specific page views and interactions with video content

Coursera also offers the opportunity to post optional enrollee surveys, which allows its university partners to gather specific information about individual students. UCI has opted to use Google

2. Signature Track is a fee-based Coursera option that allows students in select courses to earn a Verified Certificate for completing their courses. It offers identity verification, verified certificates, and sharable course records. For more information, please see <https://www.coursera.org/signature/>

3. Certain courses have been approved by the American Council on Education (ACE) for ACE CREDIT®. Institutions belonging to the ACE College and University Network can elect to recognize ACE credit recommendations and allow their students to apply those courses toward their formal degree or certification requirements. For additional information, please see <http://www.acenet.edu/news-room/Pages/The-ACE-CREDIT-College-and-University-Network.aspx>

Drive (a web-based shared document environment) to accumulate the results of these surveys for future analysis comparing student populations among different courses and among different offerings of the same course. (this last type of analysis is important for continuous quality improvement because it helps to evaluate the effects of changes made to a course). In the survey, we ask students to voluntarily provide their email address, which then allows us to match survey responses to student activity data available from Coursera for additional levels of research.

Here are several notable, early findings from the surveys (note that self-selection bias limits the statistical validity of the results):

- While our MOOCs generally do reach an international audience, there are noticeable geographical differences associated with course topics. For example, our most recently scheduled set of courses, which focus on education topics, appears to be attracting students mostly from North America and Europe.⁴ In contrast, the first set of courses we offered earlier this year attracted students from a wider global region.
- Among the students currently enrolled in our upcoming education courses, about half of them have never previously attempted an open education course of any kind and another 38 percent say they have taken only one or two MOOCs previously. The remaining 22 percent reported having taken three or more MOOCs. So, for the majority of this particular group of students we see that MOOCs are still a “new thing,” an experiment.
- In terms of demographics, the enrollees in the education courses tend to be older than typical undergraduate university students.⁵ Twelve percent of the enrollees are below the age of 24 with 71 percent being between 25 and 64 years old.
- Ninety-one percent of the students report holding a bachelor’s degree (or higher).

These very general findings reflect the kind of information that is available and suggest the types of research questions that we may begin asking.

4. These courses have not started, yet, but are already attracting enrollments

5. Undergraduate university students in the United States are usually between 18 and 24 years old.

Barriers to MOOC-based learning research

The huge potential for MOOC-based learning research may not be realized because of some barriers that are emerging in early attempts to capture data. These barriers are technological, legal, business-related, cultural, and organizational.

1. Capturing meaningful individual data about MOOC students is difficult. The very definition of a MOOC implies very easy access to the learning experience. The process of registering for a course should therefore not involve a lengthy process. However without knowing more about the MOOC student, the potential power of research is diminished. The optional survey scheme offered by Coursera and described above has the disadvantage of introducing a self-selection bias.
2. There are legal, ethical, and business reasons why capturing information that can lead to the identification of individual students is problematic. For instance, MOOC-supplying universities are subject to student privacy laws and, while it is unclear how these laws might apply to a non-matriculated student taking a free course, institutions and MOOC providers are understandably wary of transgressing the rules. Also, MOOC providers, including Coursera, are understandably concerned that if any individually identifiable student data were misused by a researcher the negative public relations consequences could seriously injure their business models and reputations.
3. MOOC audiences are heterogeneous and access is universal. In April 2013, at its first annual conference, Coursera reported that its course takers had an average of 16 years of education and that a high percentage of them came from overseas. We conclude from this that most of the MOOC participants probably were not matriculated US college students. However, things may begin to change in the near future as new MOOCs that bear academic credit are being designed specifically for college students.
4. Institutional rules may inhibit MOOC-related research. While Institutional Review Board (IRB) approval is not required for program evaluations, the most interesting research focuses on individual student learning which definitely does require

IRB approval. The lack of full control over the data from MOOCs, data that usually comes from third parties, may be an issue.

5. The ownership of data may also be a problem. Who owns the data? Could there be any restrictions on the publication of results that make use of those data? While Coursera and other MOOC providers have a very strong interest in making data available and having research performed on MOOC effectiveness and improvement, there are some potentially competing issues over data ownership.
6. Publication of MOOC-related research may not conform to general academic recognition standards. Will researchers who contribute to the MOOC body of knowledge be able to receive promotion and tenure credit for this work?
7. Finally, there is no coordinating body for MOOC-related research. Such a coordinating body could identify research questions, serve as a clearing house for research activities and the dissemination of results, track replication of findings, work with foundations on their particular goals for research, convene meetings of researchers (or even hold conferences), establish Internet forums and discussion groups, and perform many other coordinating functions.

These barriers are not insurmountable, particularly because of the significant interest in tracking the results that could be obtained from research and also because all of the players in the MOOC game - MOOC providers, university administrators, faculty authors, academic senates, foundations, the general public, policy makers and, most important of all, students - have a stake in the results of MOOC research.

MOOCs as disseminators of learning innovation

By their very nature, MOOCs are open and accessible to virtually anyone in the world. Consequently, innovative teaching approaches employed in a MOOC can be viewed freely and instantly by students, teachers, and researchers alike - wherever they happen to be located. An important benefit of introducing pedagogical innovations into the context of a MOOC is that those innovations

can be tested with real students in a real online learning setting. Researchers can monitor the MOOC and upon its conclusion evaluate the effectiveness of the innovations it introduced. They can then publish the results of their research, which subsequently helps other teachers evaluate and select appropriate innovative teaching techniques to employ in their own courses.

As innovations are adopted by others, the issue of research replication becomes automatic because the innovation is being used operationally in real life situations, not only in MOOCs but also in online education of all types and in many different contexts. According to this vision, the cycle of innovation (conception, testing, application, evaluation, replication, and general adoption) - absent intellectual property barriers and within a defined society of researchers and educators - becomes institutionalized and recognized as a part of the day-to-day application of instructional technology by institutions and faculty.

Conclusion

While the picture painted here of the potential role of MOOCs in education research may be a bit overplayed and the barriers dismissed too quickly, there is enough evidence right now to suggest that the potential for MOOCs to transform higher education resides in more than just lowering cost and improving access. MOOCs are only a special category of open education, which has existed for over 11 years in the form of Open Educational Resources (OER), OpenCourseWare (OCW), and the huge volume of open educational content currently available easily on the web with YouTube and iTunes. The potential to produce new knowledge in teaching and learning methods, illustrated here by a comparison to the advances in cosmology made possible by the Hubble Space Telescope, is certainly possible as is the rapid dissemination and adoption of those innovative methods. Realizing this vision will take a great deal of work and quite a bit of luck, but as with the struggle to fund and launch the Hubble, perhaps the stars will become aligned.

Sintesi

Da quando, nel luglio del 2011, due professori della Stanford University hanno presentato i primi MOOC (Massive Open Online Courses), la ricerca su questo tipo di corsi ha interpretato il termine massive con riferimento alla quantità degli iscritti, mentre poca riflessione è stata dedicata all'analisi dei dati, di grande interesse pedagogico, che possono essere raccolti sulle migliaia di iscritti dei MOOC.

Il dibattito sull'argomento oscilla tra l'entusiasmo per le possibilità generalizzate di accesso alla formazione offerte da questi corsi e la paura che essi possano intaccare la qualità del lavoro universitario; è auspicabile, al contrario, allargare le prospettive pedagogiche dei MOOC oltre il mero obiettivo di ridurre i costi e incrementare gli accessi. Il numero elevato di iscritti, infatti, permette di sperimentare l'efficacia dei metodi didattici e di raccogliere rilevazioni statistiche di maggiore respiro, rispetto a quanto può avvenire in corsi di piccole dimensioni.

I ricercatori dell'Università della California hanno analizzato i dati raccolti nell'ambito di un anno di collaborazione con Coursera, azienda leader nell'erogazione dei corsi MOOC, per sei corsi con circa 250 000 iscritti. I dati raccolti interessano molti aspetti dell'attività didattica, ma anche la distribuzione geografica degli studenti in rapporto ai diversi argomenti dei corsi, la formazione pregressa degli iscritti, le differenze di genere e la familiarità con la didattica online. Sono stati rilevati, inoltre, i principali ostacoli per la ricerca sugli studenti dei corsi MOOC, limiti legati sia alla eterogeneità degli iscritti, sia a ragioni legali di tutela della privacy che impedisce la raccolta di informazioni che identificano gli studenti. Inoltre, esistono problematiche legate alla proprietà dei dati diffusi in maniera generalizzata all'interno dei corsi, poiché non sempre le istituzioni educative consentono la loro distribuzione massiva. Infine, un nodo da sciogliere riguarda il riconoscimento del lavoro accademico sui MOOC nell'ambito della carriera universitaria; temi tutti di notevole interesse per l'avanzamento della ricerca sui MOOC al di là della percezione attualmente diffusa.