MOOCs and the Hubble Telescope: The Big Leap for Higher Education Research

Gary W. Matkin, Ph.D., Dean, University of California, Irvine
gmatkin@uci.edu

Abstract
The University of California, Irvine (UCI) has over 12 years of experience in open education and has most recently focused on MOOCs. This experience has periodically been distilled into research on human learning and specifically adult learning. This research has served to 1) expand the knowledge of learning, 2) feed into a continuous improvement cycle 3) attract research funds to the campus, and 4) serve as an integral part of a strategy to gain faculty support for online and open education.

As one of the first 32 institutional members of Coursera, UCI gained the opportunity to extend its research on learning to new levels. Funding from National Science Foundation, The William and Flora Hewlett Foundation, and internal sources was used to form and find answers to carefully posed research questions. In the process, UCI has discovered the potential of MOOC-related research and has identified some barriers to further progress.

This presentation will briefly describe UCI’s research method and findings from our MOOC experience to illustrate the very preliminary advantages that MOOCs provide researchers and why MOOCs might be as important to learning research as the Hubble Telescope has been to cosmology. It will also describe the barriers MOOC and OER researchers face, particularly in dealing with data ownership and Personally Identifiable Information, and what the OCW community might do to remove those barriers.

Keywords
MOOCs, MOOC Research, OER, UCI, Coursera

UCI’s History in Open Education
UCI’s research in MOOCs is being conducted against a long history of involvement in open education innovation. By the time the first widely recognized MOOC happened (July 2011) UCI had over ten years of experience with producing open educational resources (OER) and OpenCourseWare (OCW). Starting in late 2000, UCI received a series of grants from The William and Flora Hewlett Foundation to advance and study the OER and OCW movements. UCI was one of the first institutions to help found the OCW Consortium and is a charter member.

In addition, I served as founding treasurer for four consecutive years. Larry Cooperman, UCI’s director of open education, currently serves as the Consortium’s elected president. In November 2006, UCI started its open education initiative with the launch of its OCW website, which to date has received seven major national and international awards. The site contains 83 open courses and over 1,000 video lectures. UCI has sought to “publish” its open material widely, with material posted on Merlot, Connexions, YouTube, iTunesU, and other open sites. The UCI YouTube channel is viewed by up to 70,000 viewers per month at an average of 8.5 minutes per user. So it was a natural that UCI should become one of the first universities to join Coursera to create MOOCs. As of March 1, 2014, UCI has offered 13 MOOC courses on Coursera, producing over 624,000 enrollees. In November 2012 ACE and Coursera announced that two of the first five courses selected for ACE credit were UCI Coursera courses.

In April 2013 UCI announced the offering of 15 free undergraduate chemistry courses. This
offering consists of over 700 hours of high production quality video lectures that are published on UCI’s YouTube channel and OCW website. This collection of lectures constitutes the entire undergraduate UCI major in chemistry. UCI students and other students around the world watch our lectures for as many as 1,000,000 minutes per month. This allows UCI students to review their lectures for the class and students of chemistry from other universities to gain a new perspective on the subject.

Another “first” for UCI is the development of a MOOC based on the popular AMC TV drama “The Walking Dead.” UCI faculty, using scenes from the show, illustrate and teach introductory university-level subjects in ten modules, ranging from how mathematics can predict the spread of an epidemic, to the psychological effects of a major catastrophe, to principles of leadership in a time of crisis. This is the first time that a widely watched television show has been used as a means to engage thousands of people in a university-level MOOC.

A Typology Emerges
With this background of innovation and experimentation, UCI felt a responsibility to learn from what it was doing in this new field and to share that learning with others. Each of its projects has been subject to public reflection through presentations and publications. UCI has pushed even further on its research agenda with the sudden controversy regarding MOOCs and their future impact on higher education by engaging its faculty to perform high-quality research on the effects of MOOCs on their students and on specific research questions.

As research questions emerged, there also developed two separate but interrelated research agendas. The first category involves examining the phenomenology of MOOCs—who enrolls and why, what effects do they have on higher education, how disruptive they are, how can credit be granted for MOOCs or MOOC derivatives, what are the institutional impacts, what business models are emerging, and how well MOOCs work. The second category examines the effect of MOOCs on learning—how well do students learn, what elements are most effective in fostering learning, what services are supportive of the learning process, and how MOOCs can supplement traditional learning processes.

Examining the MOOC Phenomena
The massive nature of MOOCs has generated information about how the general public consumes free education. As an example, UCI collected data from its 11 Coursera MOOC courses that have been completed as shown in Table 1.
Table 1. UC Irvine Coursera MOOCs: Initial Log-ons.  
Source: UC Irvine Distance Learning Center, March 2014

Only about half of the students who enroll in a course actually ever log-on.

<table>
<thead>
<tr>
<th>Course Title</th>
<th>Quarter/Year</th>
<th>Registrations</th>
<th>Registrants Logging in at Least Once</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number</td>
<td>Percent of Registrations</td>
</tr>
<tr>
<td>Emerging Trends &amp; Technologies in the Virtual K-12 Classroom</td>
<td>Fall 13</td>
<td>21,389</td>
<td>10,534</td>
</tr>
<tr>
<td>Foundations of Virtual Instruction</td>
<td>Fall 13</td>
<td>20,276</td>
<td>11,673</td>
</tr>
<tr>
<td>Fundamentals of Personal Financial Planning</td>
<td>Winter 13</td>
<td>112,228</td>
<td>54,930</td>
</tr>
<tr>
<td>Intermediate Algebra</td>
<td>Winter 13</td>
<td>63,101</td>
<td>23,662</td>
</tr>
<tr>
<td>Pre-Calculus</td>
<td>Fall 13</td>
<td>48,503</td>
<td>36,178</td>
</tr>
<tr>
<td>Preparation for Introductory Biology: DNA to Organisms</td>
<td>Summer 13</td>
<td>37,934</td>
<td>15,563</td>
</tr>
<tr>
<td>Principles of Public Health</td>
<td>Winter 13</td>
<td>20,956</td>
<td>8,597</td>
</tr>
<tr>
<td>Science from Superheroes to Global Warming</td>
<td>Winter 13</td>
<td>17,242</td>
<td>5,789</td>
</tr>
<tr>
<td>The Power of Macroeconomics: Economic Principles in the Real World</td>
<td>Fall 13</td>
<td>30,550</td>
<td>21,158</td>
</tr>
<tr>
<td>Microeconomics for Managers</td>
<td>Winter 13</td>
<td>47,995</td>
<td>17,844</td>
</tr>
<tr>
<td>The Power of Microeconomics: Economic Principles in the Real World</td>
<td>Fall 13</td>
<td>24,004</td>
<td>15,358</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>444,178</td>
<td>221,286</td>
</tr>
</tbody>
</table>

This indicates that of the 444,178 participants who enrolled in UCI’s Coursera’s courses, 222,862 (50%) did not engage at all in the learning activity. The range of percent of log-ons for the 11 courses was from a low of 37.2% to a high of 74.6%.

Table 2 shows some measures of persistence. Clearly, participation in the course declines as the course progresses. For instance, while 74% watched the first video, only 25% watched at the midpoint. But the rate of videos watched by the end of the course was about 18%, which indicates higher rates of persistence on the relatively passive measure of video watching. For the more active measure, like taking an exam, the numbers were 40.6%, 15% and 11%, a similar pattern.

Table 2. Persistence Statistics for 11 Coursera Courses.  
Source: UC Irvine Distance Learning Center, March 2014

<table>
<thead>
<tr>
<th>Number of Registrants Who…</th>
<th>Number</th>
<th>Percent of Registrations</th>
<th>Percent of Initial Log-ons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Registrations</td>
<td>444,178</td>
<td>100%</td>
<td>n/a</td>
</tr>
<tr>
<td>Total Initial Log-ons</td>
<td>221,286</td>
<td>49.8%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Number of Registrants Who…</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watched First Video</td>
<td>165,862</td>
<td>37.3%</td>
<td>74.9%</td>
</tr>
<tr>
<td>Took First Quiz</td>
<td>89,769</td>
<td>20.2%</td>
<td>40.6%</td>
</tr>
<tr>
<td>Watched Video at Course Midpoint</td>
<td>54,789</td>
<td>12.3%</td>
<td>24.8%</td>
</tr>
<tr>
<td>Took Quiz at Course Midpoint</td>
<td>33,116</td>
<td>7.5%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Watched Last Video</td>
<td>39,493</td>
<td>8.9%</td>
<td>17.8%</td>
</tr>
<tr>
<td>Completed Final Exam</td>
<td>24,083</td>
<td>5.4%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Passed the Course</td>
<td>16,306</td>
<td>3.7%</td>
<td>7.4%</td>
</tr>
</tbody>
</table>
These results can be compared with data published by Reich (2014). Using data from edX courses, Reich found that of the over 800,000 students enrolled in 17 courses, 300,000 (37.5%) never participated, 450,000 (56%) viewed less than half the courses, and 43,000 (5.3%) earned a certificate.

These data lead to valid conclusions because of the large sample size and not from any traditional research or statistical method. Because the findings are from independent platforms and have similar results, the data here provide sound benchmarks by which to judge and compare other MOOCs. This kind of research is likely to advance rapidly as MOOC platforms are able to generate more data in usable form. On February 20, 2014 Harvard and MIT released “a set of open-source visualization tools for working with a rich trove of data from more than one million people enrolled in 17 of the two institutions’ massive open online courses, which are offered through their edX platform.” (Biemiller, 2014).

Two days later, Coursera issued its Quarter 1 2014 Roadmap, announcing its objective to “deploy standardized Coursera-wide survey questions and instrument time-on-site measures.” (Koller, Ng, 2014). The early version of Coursera’s release also includes visualization tools.

MOOCs and Learning about Learning
The most important contribution MOOC research can make is to help us understand how students learn and how MOOCs might support the learning process for students, particularly at-risk students.

For instance, a UCI professor of biology offered a Coursera MOOC in the summer of 2013 designed to prepare students to be successful in Biology 1. The course enrolled nearly 37,000 students and provides the conceptual framework, prior knowledge, and skills needed to be successful in a course that many students, even those from the top 10% of their high school classes, do poorly in. Since Biology 1 is a gateway course to a medical career, a poor showing in this freshman course can be devastating to students. UCI counseled many of its entering freshmen into this course and the professor has tracked the academic success of these students in the Fall 2013 quarter. Results of this research will be available in May 2014.

The following is a report from a UCI researcher involved with Math MOOCs:

UCI has a rich data set from which to explore the effect of course format on student success in math courses. This data includes student grades and final exam scores from face-to-face, online, and MOOC versions of Pre-Calculus and similar grades and final exam scores in a subsequent Calculus course. The Pre-Calculus final exam from the various course formats has been produced in different ways, but the problem types and topics covered are controlled between the various course offerings. The Calculus final exam is a common exam administered across all sections of the course and is designed to be comparable between quarters. Further, all three delivery approaches to the Pre-Calculus course utilized ALEKS® learning software, which has a very rich set of learning analytics available to study.

Some key preliminary findings on this data include:

- Students in online vs. on-ground Pre-Calculus courses performed comparably on equivalent final exams.
- Students in online courses spend more time on task in the course than on-ground
students.

- MOOC students who opt to use the paid ALEKS software in parallel with the course achieve and complete the course at a higher level than those who did not.
- There are no discernible differences in subsequent Calculus course success between students taking on-ground and online courses.

The Problems of MOOC Research

UCI’s experience in these research projects has revealed some inherent structures that create problems for researchers.

First, because of the many variables involved, research comparing one method of teaching (say, using MOOCs to introduce the subject as in the example above about the pre-biology MOOC) with another (not using MOOCs). While in the example, the teacher variable remained constant, it is often difficult to control for the teaching component, student composition, and other variables (time of day of class, degree of interaction among students, etc.). It is also hard to isolate individual elements within the experiment. For instance, in the example, if the MOOC contained an element of adaptive learning or learning analytic, it would be difficult to isolate the MOOC component (free course) from the mix of different treatments. Basically it is difficult to do rigorous, scientific research wherein a hypothesis is posed and then data is collected to prove or disprove the hypothesis. MOOC research tends to be descriptive and relational rather than attempting to attribute narrowly defined cause and effect relationships. MOOC research is characterized by the analysis of large data sets to see what relationships appear. The very large sample size (n’s) provides the validity of this approach rather than more detailed statistical methods.

Another category of MOOC research problems is more associated with actually trying to do research, gather, and obtain data. One of the main issues surrounds the care that needs to be taken over Personally Identifiable Information (PII). Students, by logic and federal law, have a right to privacy. Any information that can be directly attributed to a student (through a name, email address, student I.D. number, or pattern of course interaction) must be isolated and protected. Coursera identifies three “ tiers” of information: PII, unanonymizable data (such as peer graded assignments feedback that may contain PII submitted by someone other than the personally identified student), public forum, and general course data. Public forum data are student comments that are publically available and were voluntarily made public by the student. This includes responses students may post to a course discussion or “upvotes.” General course data are anonymizable data about student activity in a course such as time stamped responses for anonymizable assessments that can be derived from completion rates by assignment or lesson.

Once PII is released from a third party such as Coursera, (or generated by the institution itself), institutional policies must be followed to protect the institution from exposure to misuse of the data. Usually a form of human subjects protocol (IRB) needs to be competed and followed.

So the first structural element in MOOC research is the difference in how research is conducted between MOOCs offered directly by the researching institution and research offered through a third party such a Coursera. The third party has a huge stake in how the research is conducted. Any misuse of data collected in the researching process will fall on the third party, with potentially devastating effects on the reputation of the platform. This is true even if appropriate steps have been taken to protect the third party legally. The main
issue is the use of PII, which is sometimes necessary to obtain the kind of information necessary to carry out research.

There also is a data ownership issue that must be addressed. In the Coursera case, it owns the data that it collects through its registration and student tracking system. However, Coursera does allow the partner institution to conduct voluntary surveys of enrolled students. And now, starting the first quarter of 2014, Coursera will conduct pre-course surveys of students enrolling and will make that data available to all. Coursera, in following its interest in fostering research in the field, offers its data to partner institutions, but under carefully controlled circumstances. The PII issue comes into play here also—data from one partner institution is not available to other institutions except through the “data coordinators” on the campuses.

With regard to mining data from third party platforms, and using Coursera as an example, we have (among many additional possibilities) the ability to determine:

1. Percentage of students who completed each assignment and who received a certificate of accomplishment.
2. Number of videos watched and the average time each video was watched.
3. Percentage of students who watched videos.
4. Percentage of students who participated in the forums.
5. Summary grade information for individual students (not PII) and the components of that grading (peer assessment, assignment completion, forum posts).
6. Failure to engage (drop out rates) per week.
7. Summary of individual and cumulative grades on each quiz.
8. Summary of individual data for each “assignment part” including individual answers to quiz questions.

With regard to surveys, the amount of information is usually limited only by the time and effort it takes for a student to fill out the survey, with longer surveys producing lower response rates. Some data elements of such surveys might include:

1. PII information such as names, addresses, and email addresses.
2. Gender, age, level of educational attainment, native language, socio-economic status, profession, occupation (teacher, student, etc.)
3. Level of satisfaction, elements of satisfaction.
4. How MOOC learning was used and implemented.
5. Predisposition for consuming more MOOCs.
6. Prior level of knowledge of offering institution, attitude toward offering institution after MOOC.

From these data sets we could, theoretically, compare engagement levels (through assignment completions, number of words or upvotes posted, or completion of peer review assignments) among and between students from different countries. At the most granular level we might be able to determine whether assignments have a cultural bias. We could compare the level of engagement with the use of the material (use in classrooms by teachers) and see patterns that could help us adjust the course to be more useful. In fact, just with the limited data sets described above, the information that could be collected appears amazingly broad and of high interest, but the real potential of MOOC research goes well beyond this scope.
A final structural element is the difference between what might be called “classroom research” and “program research.” Classroom research is the type that allows instructors to improve their courses—seeing responses to individual questions, determining how long students spend on any particular assignment, comparing engagement measures (number of words offered, number of “likes” posted) from assignment to assignment. Program research is conducted primarily to measure the effectiveness of the overall program and what types of students it served. Examples of such research might be the geographical distribution of enrollees, this variable correlated to completion statistics, the percentage of students accessing the MOOC from tablets vs. lap tops or desktops.

**MOOC Research and the Hubble Telescope**

In 1990, NASA launched the Hubble Space Telescope into orbit, making it possible to view the universe unhindered by the distorting effects of the earth’s atmosphere. In an unprecedented move, data from the Hubble was made widely available even to amateur astronomers. Over 9,000 peer reviewed papers have resulted in significant breakthroughs in astronomy and cosmology, such as confirming that the universe is expanding at an accelerating rate and that black holes exist.

MOOCs have a similar potential for learning research. One of the systematic barriers to useful and generalizable research in learning is the problem of validation. Most learning research has been restricted to small samples of students with limited potential for generalizing beyond a narrowly defined set of variables. However, MOOCs now make it possible to examine the relative effectiveness of alternative-learning treatments across large numbers of students, providing statistical validity often lacking in smaller-scale studies. “A/B” testing whereby alternative-learning treatments can be delivered to the same population can be a source of new knowledge about how people learn and what is effective.

But there are more reasons for optimism about MOOCs. Much MOOC research is very inexpensive and the data can be easily obtained and manipulated. If the transfer of data can be resolved, many more researchers can gain the base data upon which to perform analyses, much like the Hubble telescope model.

Replication of results can be quickly performed and information about research results can be easily communicated. And, because the object of research is innovation in the field of learning, innovative techniques and processes can be disseminated rapidly around the world and put into practice.

**The Future of MOOCs**

Much has been written about MOOCs and their potential disruptive impact on higher education. Some commentators welcome this impact seeing MOOCs as a way of decreasing the cost of higher education and improving learning. Others view MOOCs as disruptive to the true intent of higher education, as the commodification of education, and as a movement away from the traditional values of university education. More likely than either of these predictions is the probability that the biggest impact of MOOCs will be an increase in world knowledge about learning and the learning process. That knowledge, like most knowledge in the world, will be derived from a scientific investigation of learning through MOOCs conducted by university researchers.
References


License and Citation
This work is licensed under the Creative Commons Attribution License http://creativecommons.org/licenses/by/3.0/. Please cite this work as: Matkin, F. (2014). MOOCs and the Hubble Telescope: The Big Leap for Higher Education Research. In Proceedings of OpenCourseWare Consortium Global 2014: Open Education for a Multicultural World.