MOOC Research: What Can We Do with Big Data?

Gary W. Matkin, Ph.D., University of California, Irvine

gmatkin@uci.edu

Abstract
MOOCs offer research objects that have the potential to address many of the issues higher education researchers face. They present new and unique opportunities to understand how people learn across a broad spectrum of educational mediums. MOOCs cross the boundaries between formal and informal learning in an unprecedented way, with each MOOC course offering opportunities for researchers to study how people select and engage with learning resources. This paper will identify important questions: how are these research efforts being focused? What are they trying to learn? What impact are they having? What are they revealing about higher education? It also will explore the current state of MOOC research, summarize the approaches being taken, highlight some of the results that are coming from the research, and make predictions about what we might expect in the future.

Keywords
MOOCs, MOOC Research, OER, UCI, Coursera

We are now in the third year of the “Massive Open Online Course (MOOC) era” which began in July 2012 with Stanford’s two artificial intelligence courses. While some see MOOCs as a temporary phenomenon, doomed by higher education’s resistance to change and lack of financially sustainable models, it’s become clear that MOOCs are here to stay. As MOOCs evolve, sponsors and courses alike are growing at a rapid rate. Yet, the effect MOOCs are having on research in higher education is lost in the hype. In fact, MOOCs have stimulated higher education research. Harvard and MIT jointly invested $30 million in edX by using opportunities for research as part of the rational for such a huge investment. Anything new excites research questions. And those willing to pay for the answers to the MOOC phenomenon include some unusual patrons such as Google, Yahoo, the Gates Foundation, and UNESCO.

MOOCs offer research objects that have the potential to address many of the issues higher education researchers face. They present new and unique opportunities to understand how people learn across a broad spectrum of educational mediums. MOOCs cross the boundaries between formal and informal learning in an unprecedented way, with each MOOC course offering opportunities for researchers to study how people select and engage with learning resources. The very large numbers (or n’s) of students in most MOOCS, coming from different countries and educational backgrounds, provide statistical validity that most learning research lacks due to highly restricted populations and narrowly focused research objects.

The MOOC phenomenon itself has researchers generating interesting hypotheses of practical impact; questions that seek to create theories rather than prove or disprove them. In a course with thousands of enrollments and hundreds of completers, research questions can be answered and administered with a high degree of validity, particularly the typical “A/B” testing. Student diversity (age, nationality, educational background) allows for the creation of cells that contain many people. The relationship between formal and informal learning can be assessed and there are institutional dynamics pushing for replication. As an example, why is our completion rate so much lower than others? Much of the MOOC experimentation and
data is itself co-developed (often between institutions, foundations, or MOOC providers) which means that there is a greater impulse to share experimental results. MOOC research has attracted funding and, more importantly, has created large-scale efforts to organize and share data.

As research questions emerge from the MOOC phenomenon, a three-part typology of research agenda also has taken shape. The first category of research has centered on the MOOC experience itself: who enrolls in MOOCs and why? The second category has to do with the impact of MOOCs on higher education: to what degree are MOOCs useful and integrated into higher education? What long-term impact will MOOCs have? Will MOOCs transform higher education? Will MOOCs impact the educational deficit of third-world countries? The third category has to do with the study of learning through MOOCs: what can MOOCs tell us about the learning process, and particularly, about informal learning?

In this last category one of the most interesting research questions is to determine how learning pathways created by instructors match the pathways students naturally take in learning something. Because MOOCs are directed largely at informal learners and their platforms can trace learner behavior at very granular levels, researchers can begin to gage the gap in learning contexts between instructors, students, and diverse groups of learners. Formal and informal learning can be studied side by side. Cross national studies of learners from different countries can help us understand how MOOCs can be effective in delivering low cost education and how they must be localized. So far, no serious research has been done to understand what MOOCs might reveal in these areas. What has been done is a few studies on whether MOOCs can help students (at risk, low income, marginally qualified) do better academically in targeted academic courses.

For instance, the University of California, Irvine, (UCI) counseled at-risk, entering freshman to take a “pre-biology” MOOC that was offered through Coursera. These students were given an incentive to complete the Coursera course “with distinction” to be able to transfer into biology as a major more quickly than if they did not take the course. The incentive worked; about 450 UCI students joined the 37,000 who enrolled in the MOOC. UCI found that those who competed the MOOC did significantly better than the two control groups, one that engaged in, but did not complete the Coursera MOOC, and one group that did not enroll in the MOOC.

Another example of the examination of learning from MOOCs comes from 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.

In addition, university studies have found that isolated underprivileged students are less likely to complete a MOOC, yet when they do they are more likely to do so with “distinction.” At risk students who had off line help with MOOCs were more successful.

MOOCs also offer the opportunity for researchers to study how systems and institutions in higher education react to potentially disruptive change. Patterns of adoption and rejection reveal where innovation actually originates, how it spreads, where, and how it persists or dies. Another related avenue for MOOC research is towards identifying the optimal learning settings. Many institutions are experimenting with using MOOC course assets in parallel with on-ground, campus-based instruction. Research with careful controls for confounding factors will allow us to investigate learning environments at a scale not previously possible in our educational system.

While higher education research can be significantly advanced through analyses of MOOCs, there are some barriers that need to be addressed. These issues include personally identifiable information (PII), ownership of data, the “messy” nature of the data, and intellectual property constraints. 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. Starting the first quarter of 2014, Coursera conducted 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.

As long as these issues are clearly understood, these barriers have a resolution. The questions, guidelines, and ethics that we have been using were designed for a different world of data production than we face now. MOOC research will push us into this new world. Data will come before theory and theory will be simultaneously developed and modified as results are tabulated. More importantly, the effects on individual learning will immediately feed back into research agendas as theories are rapidly tested in the field.

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/. Matkin, G. (2014). MOOC Research: What Can We Do with Big Data? In Proceedings of Open Education Global 2015: Innovation and Entrepreneurship.