EDUCATION RESEARCH

Rebooting MOOC Research

Improve assessment, data sharing, and experimental design

By Justin Reich

The chief executive officer of edX, Anant Agarwal, declared that Massive Open Online Courses (MOOCs) should serve as "particle accelerator for learning" (*1*). MOOCs provide new sources of data and opportunities for large-scale experiments that can advance the science of learning. In the years since MOOCs first attracted widespread attention, new lines of research have begun, but findings from these efforts have had few implications for teaching and learning. Big data sets do not, by virtue of their size, inherently possess answers to interesting questions. For MOOC research to advance the science of learning, researchers, course developers, and other stakeholders must advance the field along three trajectories: from studies of engagement to research about learning, from investigations of individual courses to comparisons across contexts, and from a reliance on post hoc analyses to greater use of multidisciplinary, experimental design.

EDUCATION

CLICKING OR LEARNING? Few MOOC studies make robust claims about student learning, and fewer claim that particular instructional moves caused improved learning. We have terabytes of data about what students clicked and very little understanding of what changed in their heads.

Consider four recent studies conducted on Udacity, Khan Academy, Google Course Builder, and edX (*2–5*). Each study addressed a correlation between measures of student success (such as test scores or course completion) and measures of student activity. All four studies operationalized activity similarly, boiling down the vast data available to a simple, person-level summary variable: number of problems attempted (Udacity), minutes on site (Khan Academy), weekly activity completion (Google), and number of "clicks" per student in the event logs (edX). The complexity of student activity (*6*) captured by these platforms was lost. Using simple comparisons or regressions, all four concluded there is a positive correlation between student activity and success.

It does not require trillions of event logs to demonstrate that effort is correlated with achievement. As these are observational findings, the causal linkages between doing more and doing better are unclear. Beyond exhorting students to be more active, there are no practical implications for course design. The next generation of MOOC research needs to adopt a wider range of research designs with greater attention to causal factors promoting student learning.

WATCHING WITHOUT LEARNING. One reason that early MOOC studies have examined engagement or completion statistics is that most MOOCs do not have assessment structures that support robust inferences about learning. MOOC researchers would, ideally, have assessment data with three characteristics. First, assessments should take place at multiple time points. Pretesting is critical in MOOCs, because heterogeneous registrants include novices and domain experts (*7*). Second, assessments should capture multiple dimensions of learning, from procedural to conceptual. Students who earn high grades on quantitative exam questions often show no growth in their conceptual understanding or expert thinking (*8*). Finally, courses should include assessments that have been validated by prior research, so comparisons can be made to other settings. Some recent MOOC studies meet these criteria and offer insights about which learners benefit most from MOOCs and which course materials may best support learning (*9*). With greater attention to assessment in course design, researchers can make stronger claims about what students learn—not just what they do.

Distinguishing between engagement and learning is particularly crucial in voluntary online learning settings, because media that provoke confusion and disequilibrium can be productive for learners (*10*). Addressing misconceptions requires addressing the uncomfortable gap between our intuitions and scientific reality. Unfortunately, learners may prefer videos that present material more simply. For instance, students use more positive language to describe instructional videos that present straightforward descriptions of phenomena, even though students learn more from media that directly address misconceptions (*11*). Course developers optimizing for engagement statistics can create pleasurable media experiences that keep students watching without necessarily learning.

RETHINK DATA SHARING. Although MOOC researchers have data from thousands of students, few have data from many courses. Student privacy regulations, data protection concerns, and a tendency to hoard data conspire to curtail data sharing. As a result, researchers can examine variation between students but cannot make robust inferences about cross-course differences.

For example, Nesterko *et al.* found a modest positive correlation between frequent, intermediate due dates and MOOC completion rates (*12*). But the 10 courses they examined differed not only by their use of due dates but also by their enrollment size, subject matter, and other dimensions. Data from

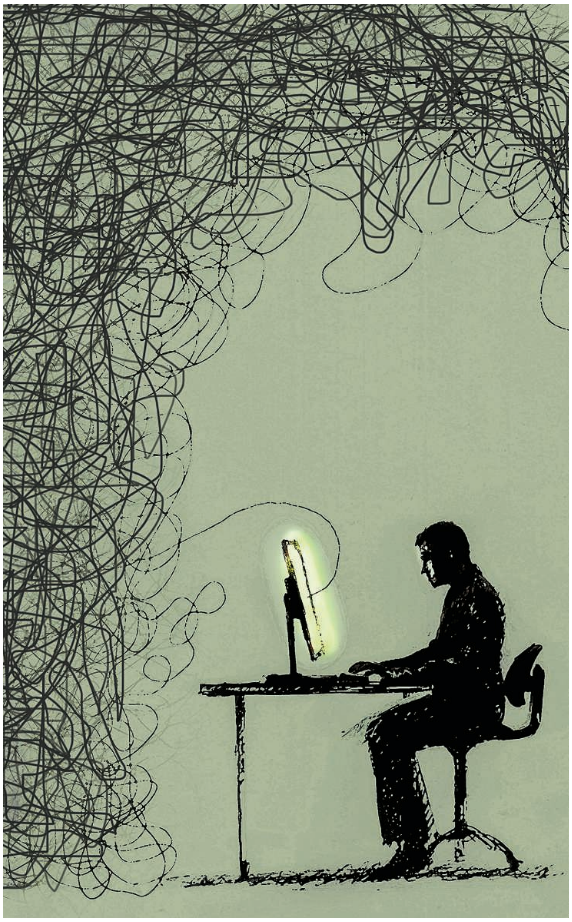

ILLUSTRATION: GARY WATERS/GETTYIMAGES.COM

HarvardX, Harvard University, Cambridge, MA 02476, USA. E-mail: justin_reich@harvard.edu

hundreds of courses will be necessary to conduct meaningful post hoc comparisons of instructional approaches.

Sharing learner data is no simple matter. Recent efforts to de-identify student data so as to meet privacy requirements demonstrate that the blurring and scrubbing required to protect student anonymity deform data to the point where they are no longer reliable for many forms of scientific inquiry (*13*). Enabling a shared science of MOOCs based on open data will require substantial policy changes and new technical innovations in social science data sharing. One policy approach would be to decouple privacy protections from efforts to maintain anonymity, which would allow researchers to share identifiable data in exchange for greater oversight of their data protection regimes. Technical solutions could include regimes based on differential privacy, where institutions would keep student data in a standardized format that allows researchers to query repositories, returning only aggregated results.

BEYOND A/B TESTS. In the absence of shared cross-course data, experimental designs will be central to investigating the efficacy of particular instructional approaches. From the earliest MOOC courses, researchers have implemented "A/B tests" and other experimental designs (*14*, *15*). These methods are poised to expand as MOOC platforms incorporate authoring tools for randomized assignment of course content.

The most common MOOC experimental interventions have been domain-independent "plug-in" experiments. In one study, students earned virtual "badges" for active participation in a discussion forum (*16*). Students randomly received different badge display conditions, some of which caused more forum activity than others. This experiment took place in a Machine Learning class, but it could have been conducted in American Literature or Biology. These domain-independent experiments, often inspired by psychology or behavioral economics, are widely under way in the field. HarvardX, for instance, has recently offered courses with embedded experiments that activate social supports and commitment devices and cause manipulations to increase perceptions of instructor rapport.

The signature advantage of plug-in experiments is that successful interventions to boost motivation, memorization, or other common facets of learning can be adapted to diverse settings. This universality is also a limitation: These studies cannot advance the science of disciplinary learning. They cannot identify how best to address a particular misconception or optimize a specific learning sequence. Boosting motivation in well-designed courses is good, but if a MOOC's overall pedagogical approach is misguided, then plug-in experiments can accelerate participation in ineffective practices. Discipline-based education research to understand domain-specific learning in MOOCs may be prerequisite to effectively leveraging domain-independent research.

There are fewer examples of domain-specific experiments that are "baked-in" to the architecture of MOOCs. Fisher randomly assigned students in his Copyright course to one of two curricula—one based on U.S. case law, the other on global copyright issues—to experimentally assess these approaches (*17*). He used final exam scores, student surveys, and teaching assistant feedback to evaluate the curricula and concluded that deep examination of a single copyright regime served students better than a survey of global approaches, providing actionable findings for open online legal education.

The next generation of MOOC research needs ... a wider range of research designs with greater attention to ... factors promoting student learning

Both domain-specific and domain-independent experiments will be important as MOOC research matures, but domain-specific endeavors may require more intentional nurturing. Plug-in experiments fit more easily in the siloed structures of academia, where psychologists and economists can generate interventions to be incorporated in courses developed by others. Domain-specific research requires multidisciplinary teams—content experts, assessment experts, and instructional designers—that are often called for in educational research (*18*) but remain elusive. More-complex MOOC research will require greater institutional support from universities and funding agencies to prosper.

RAISING THE BAR. In a new field, it is appropriate to focus on proof-of-concept demonstrations. For the first MOOC courses, getting basic course materials accessible to millions was an achievement. For the first MOOC researchers, getting data cleaned for any analysis was an achievement. In early efforts, following the path of least resistance to produce results is a wise strategy, but it runs the risk of creating path dependencies.

Using engagement data rather than waiting for learning data, using data from individual courses rather than waiting for shared data, and using simple plug-in experiments versus more complex design research are all sensible design decisions for a young field. Advancing the field, however, will require that researchers tackle obstacles elided by early studies.

These challenges cannot be addressed solely by individual researchers. Improving MOOC research will require collective action from universities, funding agencies, journal editors, conference organizers, and course developers. At many universities that produce MOOCs, there are more faculty eager to teach courses than there are resources to support course production. Universities should prioritize courses that will be designed from the outset to address fundamental questions about teaching and learning in a field. Journal editors and conference organizers should prioritize publication of work conducted jointly across institutions, examining learning outcomes rather than engagement outcomes, and favoring design research and experimental designs over post hoc analyses. Funding agencies should share these priorities, while supporting initiatives—such as new technologies and policies for data sharing—that have potential to transform open science in education and beyond. ■

REFERENCES

1. P. Stokes, *Inside Higher Ed* (2013); bit.ly/13deToN.
2. E. D. Collins, "SJSU plus augmented online learning environment: Pilot project report" (The Research & Planning Group for California Community Colleges, Sacramento, CA, 2013).
3. R. Murphy, L. Gallagher, A. Krumm, J. Mislevy, A. Hafter, "Research on the use of Khan Academy in schools" (SRI Education, Menlo Park, CA, 2014).
4. J. Wilkowski, A. Deutsch, D. M. Russell, in *Proceedings of the ACM Conference on Learning@Scale 2014*, Atlanta, GA, 4 and 5 March 2014 (ACM, New York, 2014), pp. 3–10.
5. J. Reich *et al.*, "HeroesX: The Ancient Greek Hero: Spring 2013 Course Report" (Working paper no. 3, Harvard–HarvardX, Cambridge, MA, 2014).
6. R. S. Siegler, K. Crowley, *Am. Psychol.* **46**, 606 (1991).
7. E. J. Emanuel, *Nature* **503**, 342 (2013).
8. A. Van Heuvelen, *Am. J. Phys.* **59**, 891 (1991).
9. K. Colvin *et al, IRRODL* **15**, no. 4 (2014).
10. S. D'Mello, B. Lehman, R. Pekrun, A. Graesser, *Learn. Instr.* **29**, 153 (2014).
11. D. A. Muller *et al.*, *Sci. Educ.* **92**, 278 (2008).
12. S. Nesterko *et al.*, in *Proceedings of the ACM Conference on Learning@Scale 2014*, Atlanta, GA, 4 and 5 March 2014 (ACM, New York, 2014), pp. 193–194.
13. J. P. Daries *et al.*, *Commun. ACM* **57**, 56 (2014).
14. R. F. Kizilcec, E. Schneider, G. Cohen, D. McFarland, *eLearning Pap.* **37**, 13 (2014).
15. D. Coetzee, A. Fox, M. A. Hearst, B. Hartmann, in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Baltimore, MD, 15 to 19 February 2014 (ACM, New York, 2014), pp. 1176–1187.
16. A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, in *Proceedings of the 2014 International World Wide Web Conference*, Seoul, Korea, 7 to 11 April 2014 (ACM, New York, 2014), pp. 687–698.
17. W. W. Fisher, "HLS1X: CopyrightX" (Working paper no. 5, Harvard–HarvardX, Cambridge, MA, 2014).
18. G. Siemens, *J. Learn. Analyt.* **1**, 3 (2014).

10.1126/science.1261627