Chapter 18: Diverse Big Data and Randomized Field Experiments in MOOCs

René F. Kizilcec,[1] Christopher Brooks[2]

[1]Department of Communication, Stanford University, USA
[2]School of Information, University of Michigan, USA

ABSTRACT

A new mechanism for delivering educational content at large scale, massive open online courses (MOOCs), has attracted millions of learners worldwide. Following a concise account of the recent history of MOOCs, this chapter focuses on their potential as a research instrument. We identify two critical affordances of this environment for advancing research on learning analytics and the science of learning more broadly. The first affordance is the availability of diverse big data in education. Research with heterogeneous samples of learners can advance a more inclusive science of learning, one that better accounts for people from traditionally underrepresented demographic and sociocultural groups in more narrowly obtained educational datasets. The second affordance is the ability to conduct large-scale field experiments at minimal cost. Researchers can quickly evaluate multiple theory-based interventions and draw causal conclusions about their efficacy in an authentic learning environment. Together, diverse big data and experimentation provide evidence on "what works for whom" that can extend theories to account for individual differences and support efforts to effectively target materials and support structures in online learning environments.

Keywords: Research methodology, big diverse data, randomized field experiments, inclusive science, MOOC

Massive open online courses (MOOCs) are a technological innovation for providing low-cost educational experiences to a worldwide audience. In 2012, some of the first MOOCs attracted hundreds of thousands of people from countries around the world (Waldrop, 2013), providing momentum for a disruption in higher education. Just a few years later, hundreds of institutions worldwide began to offer MOOCs on online learning platforms such as Coursera, EdX, and FutureLearn. Beyond expanding access to higher education, MOOCs have generated unprecedented amounts of educational data that have fuelled scholarship in existing academic communities and sparked interests in disciplines that were historically less involved in the learning sciences. This has amplified research in existing interdisciplinary communities and given rise to entirely new communities at the intersections of education, computer science, human factors, and statistics. For the field of learning analytics, we highlight two novel features of MOOCs that can enhance next-generation research: the availability of not just big but also diverse learner-level educational data, and the opportunity to run large online field experiments at low cost.

The first feature of MOOCs that supports innovative research is the amount and nature of the data that can be collected. MOOCs collect learner data that is deep as well as broad: fine-grained records from individual learners' interactions with content in the learning environment for a large number of learners (Thille et al., 2014). The dimensions of this recently available data enable applications of machine learning and data mining techniques that were previously infeasible. Beyond the large scale, however, the learner population is also considerably more diverse in MOOCs than in typical college courses. MOOCs attract more learners from countries that are not Western, Educated, Industrialized, Rich, and Democratic (WEIRD), the population on which most experimental social science is based (Henrich, Heine, & Norenzayan, 2010). Diverse data is critical for advancing inclusive scientific

theories and educational practices, ones that apply to a broader population. Moreover, big diverse data enables research that identifies individual differences between demographic and sociocultural groups (e.g., heterogeneous effects of an intervention) that could not be investigated in existing research with small or homogenous samples.

The second feature of MOOCs that promises to increase the pace and impact of educational research is the ability to conduct online experiments rapidly, economically, and with high fidelity (Reich, 2015). In the technology industry, this is referred to as *A/B testing* to convey that individuals are randomly assigned to one of two experimental conditions. Online experimentation enables rapid iteration to test theories and practices, because multiple experiments can run in parallel and researchers can add, delete, and modify experiments in real time and at low cost. For instance, a researcher might compare different versions of lecture videos in a course (e.g., varying the introduction to a topic and how concepts are presented) and observe performance on subsequent assessments. Once enough data has been collected, the researcher may drop those lecture versions associated with the lowest scores and add new versions based on theory and the results of existing versions, and continue iterating. In this process, the researcher may find that a particular version works best for a specific group of learners, for instance, less educated learners. This may provide novel theoretical insights and it calls for adaptive presentation of content to optimize for learning. Discovery of individual differences and responsive adaptation of content is feasible in digital learning environments with large heterogeneous samples of learners, such as in MOOCs.

The goal of this chapter is to map out these two features of MOOCs, in light of the development of the field of learning analytics, and discuss how these features might advance the theory and practice of learning and instruction. We begin this chapter with a brief historical overview of the genesis and development of MOOC initiatives. We discuss the advantages of big data and diverse learner samples for research and we review work that has begun to leverage these affordances. Then we turn to the opportunities that open up through experimentation and rapid iteration, and discuss how these have been used thus far in MOOC platforms. We close the chapter with a discussion of current limitations and ways to leverage the opportunities of large-scale digital learning environments more effectively.

THE PAST AND PRESENT OF MOOCS

The development of MOOCs occurred in the context of a long tradition of efforts to increase access to education, including distance learning (e.g., correspondence schools, radio instruction), open access universities, and open educational resources (Simonson, Smaldino, Albright, & Zvacek, 2011). However, the notion of what constitutes a MOOC fundamentally shifted between 2008, when George Siemens and Stephen Downes facilitated the first MOOC (Siemens, 2013), and 2012, when the *New York Times* declared "the year of the MOOC" (Pappano, 2012). This shift led Siemens (2013) to distinguish the original cMOOCs, which are based on the connectivist pedagogical model that emphasizes collective knowledge creation without imposing a rigid course structure, from later xMOOCs (i.e., the MOOCs of 2012 and onwards), which are mostly based on the instructionist model of lecture-based courses with assessments and a rigid course structure. Stanford University Professors Sebastian Thrun, Daphne Koller, and Andrew Ng, who re-envisioned MOOCs as digital amplifications of their lecture classes to reach a broader audience, sparked this ideological shift. This vision led to the creation of several corporate and non-profit MOOC-providing organizations, most notably Coursera, Udacity, EdX, and FutureLearn. Institutions of higher education worldwide rushed to contribute to the growing number of courses, with each course attracting tens of thousands of learners (Waldrop, 2013).

The initial excitement and momentum began to fade once it became apparent that MOOCs fell short of delivering on the promise of providing universal low-cost higher education. The first sobering piece of evidence was that only a small percentage of learners who start a course go on to complete it (Clow, 2013; Breslow et al., 2013), and although completion is not everyone's goal (Kizilcec & Schneider, 2015; Kizilcec, Piech, & Schneider, 2013), this pattern suggests that critical barriers have remained unaddressed. The second sobering realization concerned the promise of advancing access for historically underserved populations. Many MOOC learners are already highly educated (Emanuel, 2013). Moreover, learners in the United States tend to live in more affluent areas and individuals with greater socioeconomic resources are more likely to earn a certificate (Hansen & Reich, 2015). Further evidence shows that socioeconomic achievement gaps in MOOCs occur worldwide in terms of education levels and levels of national development (Kizilcec, Saltarelli, Reich, & Cohen, 2017), and, moreover, that women underperform relative to men (Kizilcec & Halawa, 2015). These patterns may be partly due to structural, cultural, and educational barriers (e.g., Internet access, prior knowledge, language skills, culture-specific teaching methods). Additionally, learners may face social psychological barriers, such as the fear of being seen as less capable because of their social group (i.e., social identity

threat) and feeling uncertain about their belonging in MOOCs from elite Western institutions (Kizilcec et al., 2017; Steele, Spencer, & Joshua, 2002; Walton & Cohen, 2007). At least in terms of completion rates, North American MOOCs have disproportionately benefited more privileged learners, posing a critical challenge for a technology designed to promote educational equity. This highlights the amplifying power of technology, namely that new technologies tend to reflect existing inequalities unless active steps are taken to address them. In fact, as evidence of a lack of support for underserved populations emerged, platform providers broadened their initial focus on prominent US partners and started pursuing international university partners, NGOs, and foreign governments. While early MOOC platforms focused on providing desktop-based learning experiences, platform development efforts also shifted towards expanding support for mobile devices as a way to increase access in developing countries, where mobile Internet is pervasive.

The issue of accreditation and certification for MOOC learning activities has been under constant consideration as the movement has matured. While content was originally provided free of charge, the value of a certificate (and a "verified certificate," where an individual's identity is linked more closely to their course activity) has attracted learners willing to pay for access to content or just to receive a certificate at the end. The certificate credential has also evolved over time. A number of institutions offer degrees independent of academic institutions (e.g., Udacity Nanodegrees); use MOOCs as a gateway opportunity to complete liberal arts courses online (e.g., the Arizona State University and EdX Freshman Academy partnership); create compact online postgraduate programs (e.g., MIT Microdegrees) with the option to use this credential as a pathway to a traditional graduate degree; and offer full graduate programs online (e.g., University of Illinois' iMBA and Data Science programs on the Coursera platform). There is an increasing supply of courses and short programs from various institutions, especially on popular topics like data science. Going forward, as more efficient marketplaces develop that better connect employers and educational institutions, we expect to see increased competition between educational institutions both to attract learners to their courses and to offer courses that can demonstrate superior workplace performance and career opportunities.

ENRICHING THEORY AND PRACTICE WITH DIVERSE BIG DATA IN EDUCATION

Theories of learning and instruction describe parts of a complex system (Mitchell, 2009). Any research that examines an instructional method or a learning strategy is therefore limited by its particular context,

for example participants' prior knowledge or the subject area. Which of the hundreds of potential contextual attributes matter in a particular instance is difficult to predict and infeasible to test. We therefore rely on scientific theory to constrain this complexity and identify the variables that matter (Koedinger, Booth, & Klahr, 2013). Nonetheless, educational theory is never conclusive or all encompassing. Empirical research in education, and the social sciences at large, tends to focus on specific contexts to reduce complexity at the expense of external validity. In particular, much empirical research in the social sciences is based on studies of people in WEIRD contexts, such as US college students who participate in psychology lab studies (Henrich, Heine, & Norenzayan, 2010). This raises questions about the generalizability of existing results and models to different contexts and populations. These concerns have also been raised specifically about research on technology-enhanced education (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014; Blanchard, 2012). To address this challenge, researchers require access to learner samples that are larger and more diverse than traditionally available. Such diverse learner samples are commonplace in MOOCs.

The supply and range of courses available on MOOC platforms has steadily increased since their initial offerings (Shah, 2015). These courses have been created by institutions around the world, including universities, museums, and national institutes. In early 2016, Coursera announced that they had reached 18 million learners worldwide.[1] Most learners are located in the United States, China, India, and Brazil, with the strongest growth in enrollment in Mexico, Colombia, Brazil, and Russia. Four in ten learners are women on average, but the gender ratio ranges from 22% in Nigeria to 55% in the Philippines. Likewise, interest in various course topics varies by gender and location: business courses are most popular in France, while Polish learners prefer computer science, the subject area with (globally) the least gender balance. Coursera learners tend to be well educated: around 80% had already earned a bachelor's degree, according to a survey in 2015 (Zhenghao et al., 2015). This pattern resembles that of FutureLearn, a MOOC platform based in the United Kingdom, used by 3 million learners in 2016[2]: 73% hold degrees and, in contrast to Coursera, 62% are women. EdX[3] and Udacity,[4] two other major MOOC providers, had served six million and two million learners by 2016, respectively. Many other institutions offer MOOCs, either through traditional learning management systems (e.g., the Canvas Network), via institutionally deployed open-source platforms (e.g.,

[1] https://blog.coursera.org/post/142363925112
[2] https://about.futurelearn.com/press-releases/future-learn-has-3-million-learners/
[3] http://blog.edx.org/edx-year-in-review?track=blog
[4] https://www.udacity.com/success

Open EdX), or through proprietary or custom developed platforms. Together, according to data collected by Class Central (Shah, 2015), 550 institutions have created 4,200 courses spanning virtually all disciplines that are reaching a remarkably heterogeneous population of over 35 million people worldwide.

While the data collected within a MOOC is high in velocity and volume, two of the three characteristics of *big data* (Laney, 2001), it can be limited with respect to variety, unless active measures are taken to achieve variety. Traditional educational systems collect both detailed demographic information (e.g., gender, ethnicity, proxies for socioeconomic status) and prior knowledge measures (e.g., prior college enrollments, high school grades, standardized test scores). However, these variables are not collected automatically in MOOCs in order to maintain a low barrier to entry. Many MOOC-providing institutions thus began to collect this data through optional surveys. A self-selected group of learners, who tend to be more committed to completing the course, are also those who tend to complete these surveys. Reich (2014) suggested that roughly one quarter of enrolled learners fill out a course survey. The sheer volume of collected survey responses can be high (often tens of thousands), though it is important to remember that these data represent a skewed sample of generally more motivated learners. Improved mechanisms for data collection that overcome current limitations for unobtrusively obtaining comprehensive information on learner backgrounds are needed. Nevertheless, currently available survey data supports the assumption that MOOC learners constitute a relatively heterogeneous population from around the globe.

Access to a heterogeneous learner population brings two major advantages for advancing educational theory and practice. First, when evaluating an instructional method or analytic model on a heterogeneous sample, the results are more representative of a diverse set of learners, which reduces the likelihood of drawing conclusions that have adverse consequences for underrepresented groups. Extending theory based on evidence from heterogeneous samples also promotes the development of more inclusive environments that support learners of various backgrounds. The second major advantage of heterogeneous learner samples is that they can reveal individual differences. Diversity is an essential ingredient for advancing an understanding of what works for whom and why – insights that enable effective tailoring of course materials and instructional methods. There is substantial room for improvement beyond tailoring to the "average learner," who may not even resemble any of the actual learners (Rose, 2016). In fact, learning scientists have identified numerous variables that influence the efficacy of instructional methods, including prior knowledge, cognitive control, mental ability, and personality (Jonassen & Grabowski, 1993). For example, prior knowledge is a well-documented individual difference (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010), such that instructional methods that are relatively effective for novice learners can become ineffective, even counterproductive, for learners with increasing domain knowledge – a phenomenon known as expertise reversal (Kalyuga, Ayres, Chandler, & Sweller, 2003). Taken together, diverse big data can advance a more inclusive science that moves beyond tailoring to averages.

Although researchers have examined countless individual differences, there is a scarcity of replication studies in the field of education. Replications account for only 0.13% of published papers in the 100 major journals (Makel & Plucker, 2014), and many of these studies rely on relatively homogenous student populations in WEIRD countries. Even if study samples were more diverse, the meta-analysis of individual studies across different learning contexts would be complicated by variation in instructional conditions, much of which remains unobserved and therefore unaccounted for (Gašević, Dawson, Rogers, & Gašević, 2016). MOOCs and online learning environments more generally can begin to address this pressing issue. These environments are particularly well suited for conducting large-scale studies with diverse samples in an authentic learning context, and it is substantially faster and cheaper to conduct an exact replication study in a MOOC by rerunning the same course or by embedding the same study in another course. In fact, MOOCs are especially suitable for conducting disciplinary research into what instructional approaches are most effective for different groups of learners. For example, in physics education, different approaches to teaching the second law of thermodynamics have been proposed and tested (e.g., Cochran & Heron, 2006), but it is unclear which approach is most effective for learners from different parts of the world.

A critical dimension of diversity in MOOCs is geography, and thus culture. MOOCs assemble learners from Western and Eastern countries with their distinct cultural foundations of learning (Li, 2012). In Eastern countries (e.g., China, Japan), learning tends to be viewed as a virtuous, life-long process of self-perfection, based on Confucian influences, whereas in Western countries (e.g., US, Canada), learning is seen as a form of inquiry that serves the goal of understanding the world around us, based on Socratic and Baconian influences. Indeed, students from Confucian Asia who attend Western universities go through a period of academic adjustment (Rienties & Tempelaar, 2013) and this culture shock can hinder learning and raise feelings of alienation (Zhou, Jindal-Snape, Topping, &

Todman, 2008). This highlights the potential influence of cultural differences, and individual differences more broadly, on the efficacy of instructional approaches. MOOCs provide sufficiently diverse learner samples to investigate individual differences and refine existing theories by evaluating understudied dimensions of learner characteristics, both at an individual and group level. New insights into individual differences can inform current practices that support academic adjustment and tailored learning experiences in both in-person and online environments.

Current Research that Leverages Diverse Big Data in MOOCs

Researchers are just beginning to leverage the potential of large, heterogeneous learner data from MOOCs. Recent studies have investigated demographic and geographic differences in course navigation (Guo & Reinecke, 2014), learner motivation (Kizilcec & Schneider, 2015), persistence and achievement (DeBoer, Stump, Seaton, & Breslow, 2013; Kizilcec & Halawa, 2015; Kizilcec et al., 2013), and socioeconomic differences in course completion of learners in the United States (Hansen & Reich, 2015) and worldwide (Kizilcec et al., 2017). Notably, although learner demographics account for significant differences in course outcomes, they provide limited improvements over behavioural log data in the context of predictive modelling (Brooks, Thompson, & Teasley, 2015a; Brooks, Thompson, & Teasley, 2015b). We briefly describe two examples from the literature that highlight a range of possible approaches for leveraging diversity in MOOCs.

We first consider work by Kulkarni, Cambre, Kotturi, Bernstein, & Klemmer (2015), who set out to harness the diversity of MOOC learners to improve engagement and learning. They identified the relatively low level of social interaction in MOOCs as an opportunity for innovation and research. To address this shortcoming, they engineered a peer discussion system that puts online learners in touch with others in the course via group video chats. A series of experiments were conducted to examine the influence of group composition on performance on assessments and to refine the design of the peer discussion system. In three experiments, learners were assigned into discussion groups with high versus low geographical diversity in terms of how many countries were represented. Different outcome measures of performance were assessed in each experiment, encompassing an open-ended question to assess conceptual understanding of the session, scores on weekly "homework" assessments, and the final exam score. Confirming the authors' hypothesis, high-diversity peer discussions yielded short-term improvements in performance. Gender diversity, in contrast, showed no effect overall or differentially by diversity condition, counter to what might be predicted by prior work (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Their research demonstrates a promising avenue for leveraging diversity as an educational asset. It also begins to examine individual differences that may influence the efficacy of this approach by testing for gender effects and evaluating alternative operationalizations of diversity.

A second example from the literature concerns the optimal presentation of the instructor in lecture videos. Seeing a human face can make it easier to pay attention, but it can also be distracting. The image principle posits that showing the instructor in a video does not affect learning outcomes, because the motivational benefits of social cues are counterbalanced by additional extraneous cognitive processing (Mayer, 2001). How does this finding translate to the context of MOOCs, where motivation is a critical antecedent of persistence and achievement? Kizilcec, Bailenson, and Gomez (2015) found that 35% of MOOC learners in a course preferred watching videos without the face when given a choice, mainly because they found the face too distracting. Then, in a randomized experiment in the same MOOC, the default video that constantly showed the instructor was compared to a strategic version that omitted the instructor when it was distracting. The strategic presentation raised perceived cognitive load and social presence, but it had no overall effect on persistence or course grades. However, accounting for learning preference (i.e., whether individuals preferred learning from pictures and diagrams or from written and verbal information), there was a substantial individual difference in persistence: learners who expressed a verbal learning preference were 46% more likely to drop out of the course with the strategic than the constant presentation. This demonstrates the importance of both accounting for individual differences in practice and refining existing theories. If social cues are more distracting or motivating for different people, this insight is worth incorporating in learner models for targeted instructional design.

TESTING THEORY AND EVALUATING EDUCATIONAL PRACTICES USING ONLINE FIELD EXPERIMENTS

Compared to traditional learning management systems used in higher education, MOOCs offer a limited set of options to course designers and hardly any new features. However, the large and diverse learning community behind MOOCs provides a remarkable opportunity to learn more about learning and teaching through experimental research. Much of the initial research with MOOCs focused on the analysis of course log data collected by default (e.g., clickstreams)

and self-report measures from course surveys with relatively low response rates. The recent availability of instructor-facing experimentation features in MOOCs has enabled researchers to conduct simple randomized experiments. Here we review three streams of experimental research – one concerned with small encouragements to promote engagement and learning, another concerned with changes to the course content and structure, and a third where MOOCs serve as a laboratory for studying general phenomena – and discuss methodological considerations going forward.

Three Streams of Published Experimental Research in MOOCs

One stream of experimental research has focused on small encouragements or nudges to improve course outcomes. These types of interventions can be conducted through email, for example, by randomly assigning learners to receive different messages. A number of studies employed A/B tests to increase participation in discussion forums. Lamb, Smilack, Ho, and Reich (2015) tested three treatments (a self-test participation check, discussion priming with summaries of prior discussions, and discussion preview emails about upcoming discussion topics) and found that the participation check increased forum activity over the default control condition. Kizilcec, Schneider, Cohen, and McFarland (2014) tested framing effects in email encouragements for forum participation in two experiments and found that a collectivist framing (i.e., "learn together", "help each other") reduced participation relative to an individualistic or neutral framing. Martinez (2014) tested framing effects using a social comparison paradigm (Festinger, 1954). Learners received an email with either an upward social comparison (describing how many learners outperform you), a downward social comparison (describing how many perform worse), or a control message omitting any social comparison. While the downward comparison motivated high-performing learners, struggling learners benefited from the upward comparison. Finally, Renz, Hoffmann, Staubitz, and Meinel (2016) found that emails exhibiting popular forum discussions and unanswered questions increased forum activity, and that reminder emails about unseen lecture videos increased course activity (i.e., lecture views), compared to other reminders. However, a downside of email interventions is that researchers typically cannot observe who opened the email and was exposed to the treatment, which raises an analytic challenge for estimating treatment effects (cf. Lamb et al., 2015). Survey experiments, with the experiment embedded inside a survey, offer an alternative. One study had survey-takers randomly assigned to receive either tips about self-regulated learning or a control message about course topics, but found no improvement in

course outcomes (Kizilcec, Pérez-Sanagustín, & Maldonado, 2016). A potential downside of experiments in optional surveys is self-selection into the study, which tends to skew the sample towards more committed learners who may respond to the treatment differently from those who opted against taking the survey. In general, although small nudges can have surprisingly large effects on human behaviour (Thaler & Sunstein, 2009), most experiments in MOOCs have yielded small or non-significant results.

Another stream of experimental research has examined theory-based changes to course content and course structure. Renz, Hoffmann, Staubitz, and Meinel (2016) assessed the impact of providing learners with an "onboarding" session, an interactive tour that explains the course structure and navigation, but they found no improvements in course engagement. Following a quasi-experimental approach, Mullaney and Reich (2015) compared two consecutive instances of the same course with different content release models, staggered versus all-at-once presentation of materials. They also found no significant difference in persistence and completion rates. To facilitate two established learning strategies (retrieval practice and study planning), Davis, Chen, van der Zee, Hauff, and Houben (2016) tested weekly writing prompts that asked learners to summarize content and plan ahead. Yet, once again, no improvements in course persistence and completion were detected. Building on multimedia learning theory, Kizilcec and colleagues (2015) tested how the presentation of the instructor's face in video lectures influences attrition and achievement rates and they found heterogeneous effects on attrition, as previously described. In the context of discussion forums, Tomkin and Charlevoix (2014) tested the effect of instructor contact on various course outcomes. Their high-touch condition, which had instructors respond to forum questions and send weekly summaries, did not improve satisfaction, persistence, or completion rates, compared with a low-touch condition without instructor involvement. Coetzee, Fox, Hearst, and Hartmann (2014) evaluated the impact of adopting a reputation system in the discussion forum and found that it increased response times and the number of responses per post, but it had no effect on grades or persistence. Another study evaluated different reputation systems and found that a forum badging system that emphasized badge progress and upcoming badges increased forum activity (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Most studies in this stream of research, despite using stronger manipulations, also found no significant improvements in learning outcomes.

The third stream of experimental research leverages MOOCs as a lab environment to test general the-

ories in a real-world context. For example, to test for (potentially unconscious) bias in online classes, Baker, Dee, Evans, and John (2015) planted messages in discussion forums across 126 MOOCs (1,008 messages total, eight per course) and randomly assigned learner names to be evocative of different races and genders. They found evidence of discrimination: instructors wrote more replies for white male names than for white female, Indian, and Chinese names. In testing social-psychological barriers to achievement, Kizilcec, Saltarelli, Reich, and Cohen (2017) found that theory-based intervention activities, designed to mitigate concerns about not belonging in the course, can effectively close the global achievement gap between learners in more versus less developed countries. On the benefits of diversity, Kulkarni et al. (2015) tested the role of geographical diversity in peer video discussion and found that being in a more diverse group improved subsequent test performance. Leveraging a natural experiment in peer assessment, Rogers and Feller (2016) found that exposure to exemplary peer performance causes attrition, due to the upward social comparison that undermines motivation and expected success. Again in the context of peer assessment, Kizilcec (2016) tested how the level of transparency about the peer grading process (i.e., how grades are adjusted and computed) affects learners' trust in peer grading. Results suggest that an explanation that highlights the fairness of the procedure can promote resilience in trust for learners who received a lower than expected grade. The studies in this stream of research focus on different phenomena in the context of MOOCs and their results hold promise for enriching theory and practice.

Methodological Considerations for Randomized Field Experiments in MOOCs

From this review of published research, it stands out that many experiments did not produce significant results. This may be surprising, given that MOOCs offer relatively large sample sizes that should render even practically insignificant differences statistically significant.[5] However, MOOC data exhibits substantial levels of variance in outcome measures (e.g., persistence, grades). While statistical power, the chance of detecting a true effect in data, increases with sample size, it decreases as data becomes noisier. When researchers underestimate the level of unexplained variance, it can result in underpowered studies that yield no significant findings. Yet this variance may actually signal the presence of individual differences that warrant further examination, for example, by

testing for heterogeneous treatment effects. In general, when evaluating and reporting on experiments in MOOCs, it is advisable to focus on the magnitude of treatment effects in addition to their statistical significance. Researchers should explicitly separate planned confirmatory tests of hypotheses from ad-hoc exploratory analyses. Given the overwhelming number of possible outcomes and covariate measures in MOOC data, there is a real danger of increasing the Type I error rate (false positives) as a result of multiple testing, specification search, and researcher degrees of freedom (Gelman & Loken, 2013). To address this challenge, replication, pre-registration, and use of Bayesian alternatives to frequentist hypothesis testing (e.g., Kruschke, 2013) can help build robust scientific evidence going forward.

Despite existing within a phenomenon that is only four years old, randomized experiments in MOOCs are poised to deliver significant contributions to theory in education and related disciplines. Yet the promise of online field experiments for educational improvement has not been widely realized. Limits on the availability of real-time data and the level of access required to implement complex parallel experiments mean that most researchers are still only testing one idea at a time at the pace of new courses going live. A critical step towards rapid iteration with experiments in MOOCs is laying the groundwork for adaptive experimentation to accomplish the dual goal of learning through experimentation with the learner population and iteratively providing a better learning experience. This would provide a special opportunity for advancing scholarship in disciplinary teaching and learning. For example, instead of trying out a new way of teaching the concept of recursion and comparing test results with the previous cohort (a quasi-experimental design), multiple approaches to recursion can be taught simultaneously and their efficacy determined in short order. This would enable simultaneous tests of multiple educational theories in a domain and refining theory and practice by examining heterogeneous effects – a process that currently requires a whole community of researchers and substantial resources. Williams and colleagues (2014) proposed a first concept for adaptive experimentation in MOOCs in the form of MOOClets, which are small pieces of content that adapt based on results of ongoing experiments. Going forward, dynamic assignment to experimental conditions, for instance using a multi-armed bandit algorithm (Bather & Gittins, 1990), can enable rapid iteration over course designs, especially in combination with experimental system that supports complex and parallel designs, such as PlanOut (Bakshy, Eckles, & Bernstein, 2014). Overall, randomized field experiments in MOOCs offer researchers a novel opportunity to enrich theory and

[5] As is standard practice in the social sciences, the field of educational research has adopted the $p < 0.05$ criterion to evaluate the statistical significance of experimental results. Thus, if there is a less than 5% chance of obtaining an effect at least as extreme as in the sample data when the null hypothesis is true, then the null (e.g., equal condition means) is rejected.

practice at a fast pace.

CONCLUSION

The ability to deploy large randomized experiments rapidly to heterogeneous learner populations has the potential to disrupt the way educational research is carried out. While in the past learning theories might arise from the careful study of a small number of highly selective environments where control over conditions is difficult (e.g., higher education classroom studies in WEIRD contexts), it is now possible to deploy randomized experiments with high fidelity to tens of thousands of learners across the globe in a single course – an unprecedented opportunity in the field.

Traditional higher education research has faced two major pragmatic constraints to experimental inquiry. Perhaps the most significant constraint is the paradigm of responsibility of instruction. In the higher education classroom, the faculty member teaching the course tends to be completely responsible for the student experience. Faculty thus tend to take an equality-driven approach rather than an experimental approach, and ensure that all students within the class have equal access to support and interventions. Innovation in these circumstances tends to be through quasi-experimental methods, where learners in a given cohort or year of study are compared with those of other cohorts or years of study, introducing more confounding variables. In MOOCs, the paradigm is different, perhaps due in part to the broader constellation of actors, including institutional administration and vendor partners, who assume some responsibility for the success of learners. The culture of some of these actors around balancing risk and reward, especially within venture-capital funded enterprises where rapid prototyping and testing is the norm, has fostered more favorable attitudes towards experimental approaches to advance learner success.

A second constraint in traditional higher education research is the acceptable amount of risk and reward made available through experimentation. With hundreds of years of higher education demonstrating value to society, and tens or hundreds of thousands of dollars on the line in tuition for a given student, it is harder for researchers to make the ethical argument to engage in high-risk research. Yet in MOOC environments, the majority of learners enrol at no cost, and few are in peril of losing their livelihood over the results of a MOOC experiment gone awry.[6] This difference is reflected in institutional policy. Many institutions have strong protections for student records and

privacy because of legal obligations such as FERPA in the United States. However, the same obligations do not exist for learners (or "users") in MOOCs, which lifts some of the constraints from policies concerning experimental research with online learners. This has two important ramifications and opportunities for researchers:

1. The population of MOOC learners is different and, in many ways, much more diverse than that of traditional educational research, in terms of learner demographics (age, race, cultural background, et cetera), prior knowledge, and motivations for taking courses. This broader representation, along with the vast numbers of learners, provides an opportunity for scholars to test the generalizability of learning theories across populations and to identity learning theories that are most applicable to specific groups of learners. This can enable quantitative approaches to problems that require such large datasets; for instance, Dillahunt, Ng, Fiesta, and Wang's (2016) study of low-income populations who use MOOCs for social mobility would be difficult to study quantitatively if not for the breadth of learners enrolled in MOOCs.

2. The ability to experiment directly in the learning platform at no cost enables researchers to leverage the volume and variance of learner data for greater scientific impact. This presents an opportunity to close the feedback loop in education by promoting the integration of research, theory, and practice. Given the breadth and depth of the learner population in MOOCs, there is a real possibility to build environments that (semi-)automatically adapt to the learner based on her experience, the experiences of other learners, and the underlying platform data.

In this chapter, we have called out two affordances of research with MOOCs – the availability of diverse big data and the ability to conduct randomized field experiments with rapid iteration – which we believe will enable a more inclusive and agile science of learning. The fields of learning analytics and educational data mining, characterized largely by their heavy adoption and investigation of computational methods, are well poised to answer this call and achieve even broader impact going forward.

---

[6] Recent approaches to accepting MOOC credentials as credit in higher education, such as through the Arizona State University Global Freshman Academy and the MIT Micro-Masters programs, have begun to change the stakes for learners.

REFERENCES

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 687–698). New York: ACM.

Baker, R., Dee, T., Evans, B., & John, J. (2015). Bias in online classes: Evidence from a field experiment. Paper presented at the SREE Spring 2015 Conference, *Learning Curves: Creating and Sustaining Gains from Early Childhood through Adulthood*, 5–7 March 2015, Washington, DC, USA.

Bakshy, E., Eckles, D., & Bernstein, M. S. (2014). Designing and deploying online field experiments. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 283–292). New York: ACM.

Bather, J. A., & Gittins, J. C. (1990). Multi-armed bandit allocation indices. *Journal of the Royal Statistical Society: Series A*, 153(2), 257.

Blanchard, E. G. (2012). On the WEIRD nature of ITS/AIED conferences. In S. A. Cerri et al. (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (ITS 2012), 14–18 June 2012, Chania, Greece (pp. 280–285). Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment*, 8. http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf.

Brooks, C., Thompson, C., & Teasley, S. (2015a). A time series interaction analysis method for building predictive models of learners using log data. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 126–135). New York: ACM.

Brooks, C., Thompson, C., & Teasley, S. (2015b). Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses (MOOCs). *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 245–248). New York: ACM.

Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM. doi:10.1145/2460296.2460332

Cochran, M. J., & Heron, P. R. L. (2006). Development and assessment of research-based tutorials on heat engines and the second law of thermodynamics. *American Journal of Physics*, 74(8), 734.

Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 15–19 February 2014, Baltimore, Maryland, USA (pp. 1176–1187). New York: ACM.

Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G.-J. (2016). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-regulated learning strategies at scale. *Proceedings of the 11th European Conference on Technology Enhanced Learning* (EC-TEL 2016), 13–16 September 2016, Lyon, France (pp. 57–71). Lecture Notes in Computer Science. Springer International Publishing.

DeBoer, J., Stump, G. S., Seaton, D., & Breslow, L. (2013). Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002x. *Proceedings of the 6th Conference of MIT's Learning International Networks Consortium* (LINC 2013), 16–19 June 2013, Cambridge, Massachusetts, USA. https://tll.mit.edu/sites/default/files/library/LINC%20'13.pdf

Dillahunt, T. R., Ng, S., Fiesta, M., & Wang, Z. (2016). Do massive open online course platforms support employability? *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '16), 27 February–2 March 2016, San Francisco, CA, USA (pp. 233–244). New York: ACM.

Emanuel, E. J. (2013). Online education: MOOCs taken by educated few. *Nature*, 503(7476), 342.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations: Studies towards the Integration of the Social Sciences*, 7(2), 117–140.

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28(January), 68–84.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. *Proceedings of the 1st ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 21–30). New York: ACM.

Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, 350(6265), 1245–1248.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2–3), 61–83; discussion 83–135.

Jonassen, D. H., & Grabowski, B. L. H. (1993). *Handbook of individual differences, learning, and instruction.* Routledge.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '16), 7–12 May 2016, San Jose, CA, USA (pp. 2390–2395). New York: ACM.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, 107(3), 724–739.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 57–66). New York: ACM.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Recommending self-regulated learning strategies does not improve performance in a MOOC. *Proceedings of the 3rd ACM Conference on Learning @ Scale* (L@S 2016), 25–28 April 2016, Edinburgh, Scotland (pp. 101–104). New York: ACM.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM.

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, 355(6322), 251-252.

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. ACM *Transactions on Computer–Human Interaction*, 22(2), 1–24.

Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. *eLearning Papers*, 37, 13–22.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, 342(6161), 935–937.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573–603.

Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M. S., & Klemmer, S. R. (2015). Talkabout: Making distance matter with small groups in massive classes. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 14–18 March 2015, Vancouver, BC, Canada (pp. 1116–1128). New York: ACM.

Lamb, A., Smilack, J., Ho, A., & Reich, J. (2015). Addressing common analytic challenges to randomized experiments in MOOCs: Attrition and zero-inflation. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 21–30). New York: ACM.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 70.

Li, J. (2012). *Cultural foundations of learning: East and west*. Cambridge University Press.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316.

Martinez, I. (2014). MOOCs as a massive research laboratory. University of Virginia.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mitchell, S. D. (2009). *Unsimple truths: Science, complexity, and policy*. University of Chicago Press.

Mullaney, T., & Reich, J. (2015). Staggered versus all-at-once content release in massive open online courses: Evaluating a natural experiment. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 185–194). New York: ACM.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501.

Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Reich, J. (2014, December 8). MOOC completion and retention in the context of student intent. *EDUCAUSE Review Online*. http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

Reich, J. (2015). Rebooting MOOC research. *Science*, 347(6217), 34–35.

Renz, J., Hoffmann, D., Staubitz, T., & Meinel, C. (2016). Using A/B testing in MOOC environments. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 304–313). New York: ACM.

Rienties, B., & Tempelaar, D. (2013). The role of cultural dimensions of international and Dutch students on academic and social integration and academic performance in the Netherlands. *International Journal of Intercultural Relations*, 37(2), 188–201.

Rogers, T., & Feller, A. (2016). Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting. *Psychological Science*, 27(3), 365–374.

Rose, T. (2016). *The end of average: How we succeed in a world that values sameness*. HarperCollins.

Shah, D. (2015, December 28). MOOCs in 2015: Breaking down the numbers. *EdSurge*. https://www.edsurge.com/news/2015-12-28-moocs-in-2015-breaking-down-the-numbers

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia, & S. Marshall (Eds.), *Open educational resources: Innovation, research and practice* (pp. 5–16). Edmonton, AB: Athabasca University Press.

Simonson, M., Smaldino, S. E., Albright, M., & Zvacek, S. (2011). *Teaching and learning at a distance: Foundations of distance education*. Pearson Higher Ed.

Steele, C. M., Spencer, S. J., & Joshua, A. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379–440.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Thille, C., Schneider, E., Kizilcec, R. F., Piech, C., Halawa, S. A., & Greene, D. K. (2014). The future of data-enriched assessment. *Research & Practice in Assessment*, 9, 5–16.

Tomkin, J. H., & Charlevoix, D. (2014). Do professors matter? Using an A/B test to evaluate the impact of instructor involvement on MOOC student outcomes. *Proceedings of the 1ˢᵗ ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 71–78). New York: ACM.

Waldrop, M. M. (2013). Online learning: Campus 2.0. *Nature*, 495(7440), 160–163.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96.

Williams, J. J., Li, N., Kim, J., Whitehill, J., Maldonado, S., Pechenizkiy, M., Chu, L., & Heffernan, N. (2014). The MOOClet framework: Improving online education through experimentation and personalization of modules. Working Paper No. 2523265, Social Science Research Network. doi:10.2139/ssrn.2523265

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.

Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2015, September 22). Who's benefiting from MOOCs, and why. *Harvard Business Review*. https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why

Zhou, Y., Jindal-Snape, D., Topping, K., & Todman, J. (2008). Theoretical models of culture shock and adaptation in international students in higher education. *Studies in Higher Education*, 33(1), 63–75.