Moving Through MOOCs: Understanding the Progression of Users in Massive Open Online Courses

Laura W. Perna[1], Alan Ruby[1], Robert F. Boruch[1], Nicole Wang[1], Janie Scull[1], Seher Ahmad[1], and Chad Evans[1]

This paper reports on the progress of users through 16 Coursera courses taught by University of Pennsylvania faculty for the first time between June 2012 and July 2013. Using descriptive analyses, this study advances knowledge by considering two definitions of massive open online course (MOOC) users (registrants and starters), comparing two approaches to measuring student progress through a MOOC course (sequential versus user driven), and examining several measures of MOOC outcomes and milestones. The patterns of user progression found in this study may not describe current or future patterns given the continued evolution of MOOCs. Nonetheless, the findings provide a baseline for future studies.

Keywords: postsecondary education; retention; technology

Higher education in the United States and other nations faces numerous challenges. Leading the list are the needs to expand access and opportunity for students to participate in and benefit from higher education and reduce the cost of higher education.

Some commentators assert that massive open online courses (MOOCs) offer a new way to address these challenges. Although the characteristics of MOOCs will continue to evolve (Dodd, 2014; McGuire, 2014), MOOCs are generally understood to be full-length "courses" that can accommodate large numbers of users from across the globe ("massive"), are available at no charge to users with minimal registration requirements ("open"), and are delivered "online," over the Internet.

MOOCs build on a history of initiatives intended to exploit available technologies to expand the reach of higher education (Universities UK, 2013). Correspondence courses began providing education at a distance around the year 1840 (Lease & Brown, 2009). Dedicated to "using new and emerging technologies" to deliver higher education to students at a distance, the United Kingdom's Open University (2014) enrolled its first students in 1971. In 1992, the Sloan Commission began funding online education projects (Picciano, 2012), and between 2000 and 2007, open education and online learning expanded considerably (Universities UK, 2013).

In 2008, Stephen Downes, George Siemens, and Dave Cormier launched the connectivist MOOC (cMOOC), which assumed that knowledge is "distributed across a network of connections" and endeavored to use technology to create a learning community (Downes, 2012, p. 9). Whereas cMOOCs rely on connectivist learning theory and encourage individuals to participate as they would in an interactive seminar, the xMOOC has a traditional lecture format (Rhoads, in press). In fall 2011, Stanford's Sebastian Thrun and Peter Norvig launched the first xMOOC, Introduction to Artificial Intelligence, with 160,000 users.

To advance the provision of MOOCs, in 2012, Thrun founded Udacity, Stanford's Daphne Koller and Andrew Ng founded Coursera, and MIT and Harvard founded edX (McGuire, 2014). By January 2014, Udacity had partnerships with 16 universities and other organizations, Coursera had 107 partners, and edX had 30 partners. With their partner institutions, these three MOOC providers had offered nearly 700 xMOOCs and counted more than 8 million users worldwide (Cusack, 2014).

The Progression of Users Through MOOCS

Although impressive, the large numbers of MOOC offerings and users do not reveal whether MOOCs can address the access and cost challenges facing higher education. One set of foundational questions that must be addressed first pertains to the pattern of user progression through a MOOC. What percentage of users not only registers for a MOOC but also completes? What

[1]University of Pennsylvania, Philadelphia, PA

Educational Researcher, Vol. 43 No. 9, pp. 421–432
DOI: 10.3102/0013189X14562423

are the milestones that predict course completion? Do users treat MOOCs not as courses to be taken linearly from start to finish but instead take advantage of MOOCs' "open" feature and partake only of some course segments? Answers to these questions are required before one can understand whether and how MOOCs might expand access to higher education and provide lower-cost higher education opportunities.

This study addresses these foundational questions using descriptive analyses of data from the 16 Coursera courses taught by University of Pennsylvania (hereafter, Penn) faculty for the first time in the 1st year of Penn's xMOOC offerings (June 2012 to July 2013). By focusing on first-generation MOOCs, this study offers insights into a past but important point in time. MOOCs are a new phenomenon with evolving purposes, priorities, processes, and approaches. We expect the pattern of progression through a MOOC to change as participant characteristics and instructional approaches shift and the population of MOOC providers and offerings increases and diversifies. Like other "early adopters" (Rogers, 1962), first-generation MOOC users are likely to have more education, higher socioeconomic status, and more financial resources than later users. MOOC offerings are also likely to change over time, incorporating pedagogical insights gained from initial teaching experiences (Grainger, 2013).

The patterns of progression found in this study may not describe current or future patterns given the continued evolution of MOOCs. Nonetheless, the findings provide a necessary baseline for future studies. The growth of MOOCs has exceeded the production of available research. As such, there are not yet shared understandings of the most appropriate definition of a MOOC "student" or the best measures of user outcomes. The definitions and measures used in this study are a point of departure for future studies.

Literature Review

Researchers have examined various aspects of participation in online and distance-learning courses, including the characteristics of online learners (Carnoy et al., 2012; Noel-Levitz, 2011), predictors of retention and persistence in online programs (Carnoy et al., 2012; Hart, 2012; Hills, 2010; Jaggars, Edgecombe, & Stacey, 2013), enrollment and persistence of working adult online learners (Council for Adult and Experiential Learning, 2007), and scalability of online initiatives (Moloney & Oakley, 2010). A few studies use randomized controlled trials to test the effects of different approaches to online learning (e.g., Bowen, Chingos, Lack, & Nygren, 2012).

The applicability of findings from online and distance-learning courses to MOOCs is unclear, given the many differences in course structure and goals. Unlike MOOCs, conventional online and distance-learning courses are not intended to attract "massive" numbers of students or be "open" to users moving in and out of a course multiple times. MOOCs may be characterized as less uniform, regulated, and centralized than other large-scale online efforts, like China's Open University or Turkey's Anadolu University.

Several conclusions may be drawn from emerging research on MOOCs. First, although attracting a large number of registrants, MOOCs are typically characterized by very low completion rates. Definitions of completion vary, with disagreement over the composition of both the numerator and the denominator. Regardless of definition, reported completion rates generally range between 5% and 12% of registrants (Cusack, 2014; Koller, Ng, Chuong, & Chen, 2013; Ho et al., 2014; Universities UK, 2013). In an examination of 17 MOOCs offered by MIT and Harvard in 2012 and 2013, Ho et al. (2014) found that 5% of the over 840,000 registrants received certificates of completion; one third did not access a course. Only 12% of the nearly 310,000 registrants in six MOOCS of short duration (i.e., 5 to 7 weeks) introduced by the University of Edinburgh in 2013 submitted assignments at Week 5 (MOOCs @ Edinburgh Group, 2013).

Available research provides limited understanding of user progress through a MOOC beyond the first and last events: registration and completion. Potential intermediary outcomes include viewing a lecture and posting to a discussion forum (MOOCS @ Edinburgh, 2013). In their study of Harvard and MIT courses, Ho et al. (2014) consider whether a user posts to a discussion forum as well as the numbers of "clicks" (i.e., "the number of discrete actions a user takes in a course"; p. 26) and "active days" (i.e., "number of discrete days, demarcated in UTC time that a user engages in some type of course activity"; p. 26). Their study finds higher scores on these measures for users who "explored" or "completed" a course than for those who only registered.

User outcomes vary across courses (e.g., MOOCs @ Edinburgh, 2013). Yet, because of the many differences in the characteristics of available MOOCs, research to date provides few insights into how course characteristics contribute to variations in user outcomes. Ho et al. (2014) observed that the 17 MOOCs in their study differed in terms of content, design, duration, instructor philosophy, learner expectations, video design and distribution, assessment, and criteria for certification. Even in four courses of similar duration (5 to 7 weeks), Grainger (2013) identified variations in course content as well as the number and length of videos, assessment type, number of teaching staff and teaching assistants, forum moderation, and use of social media.

Conceptual Model

Establishing how students progress through an educational system is a prerequisite for educational planning and decision making (King, 1972). Modelers of education systems assume that to achieve any educational outcome, students must complete a series of steps. Examining the United Kingdom's educational system, Stone (1970) and King (1972), for instance, built "conceptually and algebraically simple" models that represent educational systems as "a set of branches through which students can flow" (Stone, 1970, p. 279). These models are designed to document the "movement of individuals between different parts of the system" (King, 1972, p. 53) and predict the likelihood that an individual will progress from one step to another. These predictions allow estimates of the likely human capital yield from a cohort entering an educational process and provide a baseline for exploring the events that lead some participants to become inactive and others to complete a course or program. Such information can be used to design courses and programs in ways that improve student outcomes.

This interest in documenting the steps that predict students' progress toward various educational outcomes underlies numerous conceptual studies and reports. For instance, in their seminal piece, Hossler and Gallagher (1987) identify three stages that culminate in a student's choice of college to attend: becoming predisposed to attend college, searching for information about college, and selecting a college or university in which to enroll. Perna and Thomas (2008) frame college student success as a longitudinal process that involves several key steps: becoming academically ready for college and aspiring to attend, choosing a college to attend and enrolling, making satisfactory academic progress in college course work and persisting to degree completion, and transitioning from college to enrollment in postbaccalaureate programs and careers. Along the same lines, Moore, Shulock, and Offenstein (2009) focus on identifying "milestones" of academic progress for community college students. They define milestones as the "intermediate educational achievements that students reach along the path to degree completion" (Moore et al., 2009, p. i). They conclude that for community college students, relevant intermediate measures of academic progress focus on remedial and gateway courses and accumulation of course credits.

Identifying intermediary outcomes and milestones is a prerequisite for understanding the forces that promote and limit a student's progress toward completion of a course or degree program and identifying policies and practices that may improve desired outcomes. Because of the newness of MOOCs and competition between providers, there is not yet agreement on the most appropriate measures of academic progress through a MOOC. For instance, it is not known whether users' progress through a MOOC should be measured in a sequential fashion (defined as participating in course modules in the order specified by the instructor) or in a way that captures the actions of users who take advantage of the "open" nature of MOOCs (and thus participate in a more sporadic or nonlinear fashion). Available research also does not reveal the intermediary outcomes that predict completion of a MOOC.

Research Questions and Methods

This study addresses these knowledge gaps by examining the following research questions:

1. Do MOOC users progress through a course sequentially in the order identified by the course instructor, or do users determine their own approach to accessing content?
2. What are the milestones that predict course completion?

This study addresses the research questions using data from the population of 16 first-generation MOOCs taught for the first time by Penn faculty in the 1st year of Penn-Coursera offerings. Courses were offered and completed between June 2012 and July 2013. We obtained each instructor's permission to analyze his or her course data and approval for the study from Penn's institutional review board.

We reviewed course syllabi to identify salient course characteristics. We then conducted descriptive analyses of data generated in the courses and provided by Coursera. Descriptive analyses are a necessary first step in this new area of inquiry. Before engaging in more sophisticated analyses, the field requires greater understanding of the most appropriate definitions of a MOOC user and the best indicators of user outcomes and progression.

Determining Course Characteristics

To identify salient course characteristics, we collected and reviewed the syllabus for each course. Two members of the research team independently coded ingredients of syllabi. All members of the research team provided input and feedback, drawing from the literature review and insights from enrollment in contemporary versions of several first-generation courses and adjusting codes as necessary.

Using these procedures, we classified seven of the 16 courses as related to health, three as math and economics, three as business, and three as humanities. The 16 courses appeared to target different audiences. Although not explicitly stated in the syllabi, two courses appeared oriented toward undergraduate students, whereas seven courses were occupationally oriented and seven seemed focused on personal enrichment.

Courses ranged in length from 6 to 12 weeks. Five were 6 or 7 weeks in duration, five were 8 to 9 weeks, three were 10 weeks, and three were 12 weeks. The 16 courses offered between 20 and 75 min of video per week, with videos chunked into one to 18 segments per week that were each 3 to 45 min in length. Five of the 16 courses had 40 or fewer minutes of video per session delivered in one to three chunks, and four courses had more than 60 min of video chunked into seven or more segments.

Course assessments differed in format (e.g., quizzes, homework, final exam) and frequency. Feedback mechanisms varied. Two courses offered live chat and eight offered some type of peer assessment. The courses had between one and 13 teaching assistants.[1] This range suggests differential instructional resources and investment (assuming that teaching assistants are compensated and trained) and differential potential for student engagement and individualized feedback.

Storing and Preparing Coursera's Administrative Data

Coursera provided multiple data files for each course to the information technology (IT) department of Penn's Graduate School of Education. The IT department established a secure, standalone server to store the files and then converted the files from SQL format to .txt format. All files were stored in long (versus wide) format, with multiple observations for an individual registrant in each lecture, assignment, quiz, and assessment file.[2] The research team converted the data to wide format and used SPSS, SAS, and R to conduct the analyses.

For each course, Coursera provided files with metadata and individual records. Metadata files included variables describing various course characteristics (e.g., titles of lectures and quizzes, dates lectures and quizzes were posted online, etc.).[3] The research team matched available metadata information to course syllabi and information provided directly by instructors to develop a schedule of key course dates and activities.

Coursera provided individual-level data that were distributed across multiple data files. Together these data included unique

deidentified user ID, dates of course registration and last access, dates of downloads for each lecture, dates and grades for each assignment submission, dates and grades for each quiz submission, dates and grades for final exam submissions, final grades, and local IP addresses. For each course, the research team used the user IDs to match data across files. Unique to each course, the individual IDs enabled us to match files within a course but did not allow us to match individuals across courses.

Defining MOOC "Users" and "Outcomes"

The MOOC world has not yet agreed on definitions of a MOOC "user" or measures of academic progress and outcomes. In examining persistence at the UK Open University, Ashby (2004) reports on several outcomes (e.g., completing the final assessment, passing the course) for two populations: those who register at the course start and those who register in the first third of the course. Ho et al. (2014) identified four mutually exclusive groups: those who "only registered" and did not access the course in any way, those who "only viewed" less than half of the material for any course, those who "only explored" by accessing more than half of the course material but not earning a certificate of completion, and "certified" users who earned a certificate of completion. Taking yet another approach, Grainger (2013) defined active users as those "who viewed or downloaded a lecture, attempted a quiz, registered after the MOOC start date and/or posted on MOOC forums" (p. 27). These varying approaches challenge efforts by researchers and practitioners to draw meaningful conclusions across available studies.

Data files provided by Coursera required decisions about how to define a MOOC user. For instance, the Coursera files included users who registered for courses over a wide period of time. In our analyses, we excluded users who registered before a course officially opened for registration, assuming that these users were test cases created by the instructor, teaching assistants, IT staff, and/or others affiliated with Penn or Coursera. We also excluded users who registered for a course more than 2 months after the final course segment. We assumed that these users were unable to engage with the instructors or peers about course content and were ineligible to earn a grade. The extent to which these assumptions hold likely varies across courses; our research team did not have the information needed to test the assumptions.

Because courses ended on different dates between June 2012 and July 2013, the number of months that elapsed between a course's official end and the date that we collected data varied. We set the 2-month cutoff date in order to standardize the length of time for counting registrants. Although somewhat arbitrary, we selected 2 months after the course end date based on our visual inspection of the data across courses. This visual inspection showed a relatively small and declining number of registrants between the course end date and the 2-month mark followed by a steady low rate of new registrants thereafter. Excluding those who registered before a course officially opened for registration or more than 2 months after the course ended reduced the number of registrants in the 16 courses by about 10%, from 805,408 to 710,385. We did not count the number of data entries for registrants across all files, but this number is likely in the tens of millions.

Our analyses examine two groups of users: registrants, as defined above, and a subset of this group that we call starters. Like Ashby (2004), we define starters as those who register no later than 1 week after a course start date. We hypothesize that starters will tend to behave like students enrolled in face-to-face courses, enrolling on or about the 1st day of class and moving through a course sequentially. We expect registrants, who may enroll at any time, will tend to move through a course at their own pace and sequence. With these two approaches, the analyses shed light on whether MOOC users progress through a course in a pattern that is similar to or different from students in traditional, face-to-face courses.

Using the schedule of dates and activities produced for each course (described above), we examined the following indicators of outcomes and milestones: accessed any lecture, the first lecture, and/or the last lecture; attempted any quiz, the first quiz, and/or the last quiz; and received a final grade of 80% or higher, the rough equivalent of a B average.[4]

Because of the very high drop-off between registering and accessing the first lecture, we also examined another measure of academic progress: retention rate. We defined the retention rate as the number of users who accessed a lecture in the final module divided by the number who accessed a lecture in the first module.

We also examined whether registrants and starters progress through a course sequentially, as defined by the course instructor, or whether these users take another approach to accessing content. We defined "sequential movement" as the progression of a user through a course in the sequential order identified by the instructor. This student-level, longitudinal approach identified the share of users who accessed a lecture in the first course "module" (defined as the activities composing the first segment or week of course material), the share who then accessed a lecture in the second course module, and so on through the final course module.

Focusing on the sequential approach ignores the possibility that accessing course content may have value regardless of the order or sequence in which a user engages the material. Thus, the second "user-driven" approach considered the share of users who accessed a lecture in any course module, ignoring the order or sequencing of lecture access or whether the user accessed any other lecture. This second approach recognized that a user may choose whether and when to access learning materials.

Our descriptive analyses considered the percentages of registrants and users who achieve the outcomes and milestones in each course as well as the median and range of percentages across courses. Medians are used rather than means because the statistical distributions are skewed and have outliers.

Limitations of Available Data

The data available for this study had several limitations. First, despite considerable effort, we were unable to develop a complete schedule of dates and activities for all 16 courses. We could not identify the specific lectures that corresponded to particular modules in three courses. Using unique identifiers, matching registration data to lecture downloads was relatively straightforward. But interpreting available data was difficult as the variable

names in the metadata files were typically provided without necessary contextual information. The metadata provided a lecture's title and date of posting but did not indicate the order of the lectures or dates when lectures were made available to users.

Second, we made assumptions about some incomplete data. For instance, the date a user "last accessed" a course was missing for a number of observations. We learned from Coursera that these observations represented users who registered for a course but did not access the course homepage after the course opened. We set the last access date for these users as the registration date, but these users could have accessed course material though an avenue other than the course homepage. The metadata also did not indicate the date that the course officially opened for registration. We identified this date by visually reviewing the data and assuming that registration opened on the date of the giant spike in registrants that consistently appeared across courses.

Third, although Coursera collected extensive data on each course, we, like other researchers who examined MOOCs on other platforms (e.g., Ho et al., 2014), found that some potentially important variables were not systematically collected and that some available variables provided cursory insights. Using the data available, the analyses consider only whether a user "accessed" a lecture or "attempted" a quiz and thus do not reveal whether a user actually viewed or completed a lecture or quiz. The measures used in this study also provide little insight into learning progress or outcomes.

Fourth, this study sheds little light on variations in outcomes based on course characteristics. Our efforts to explore these variations were constrained by the inconsistent information instructors provided on their syllabi and the small number of courses with any particular characteristic. Course syllabi offered varying descriptions and levels of detail about content, learning objectives, targeted level of instruction (e.g., high school, undergraduate, graduate), and other characteristics and generally did not report other potentially useful information (e.g., instructor's experience with online instruction). Coursera has adopted a nonprescriptive approach that emphasizes the freedom of individual institutions and faculty to design their own courses (Daniel, 2012). Although this nonprescriptive approach is likely attractive to faculty, it complicates efforts to identify course characteristics and understand how course characteristics contribute to variations in user outcomes.

Findings

For the 16 MOOCs in this study, there were 710,385 registrants and 541,576 starters.[5] The number of registrants per course ranged from 13,228 to 99,451, with a median of 37,909. Most registrants were starters, as starters represented 78% of registrants. The representation of starters among registrants varied considerably across courses, ranging from 53% to 95%.

Course completion rates were low regardless of whether measured as accessing the last lecture, attempting the last quiz, or receiving a final grade of at least 80%. Table 1 shows that in no course did more than 12% of registrants or starters receive a final grade of 80% or higher. Across courses, the median share of registrants who received a final grade of at least 80% was 3% according to the sequential approach and 4% in the user-driven approach. Across the 16 courses, only 5% to 18% of registrants (median of 7% with the sequential approach and 8% with the user-driven approach) and between 5% and 19% of starters (median of 8% in both approaches) accessed the last lecture. Using the sequential approach, we found that between 2% and 13% of registrants (median = 6%) and between 2% and 16% of starters (median = 6%) attempted the last quiz.

Sequential Versus User-Driven Lecture Access

Most users appear to access lectures sequentially, as the percentages of registrants and starters who accessed the first and last lectures were similar with the sequential and user-driven approaches. Table 1 shows that for registrants (a group that includes users who registered up to 2 months after a course's official end date), the minimum, maximum, and median percentages accessing the last lecture are all 1 percentage point higher with the user-driven rather than the sequential-approach. These findings suggest that a small share of users may be taking advantage of the open nature of MOOCs to determine the order in which they access course content (in this case, the last lecture).

Milestones That Predict Accessing the Last Lecture

One milestone for MOOC users is accessing at least one lecture. Table 1 shows that across the 16 courses, fewer than half of registrants (46%) and starters (43%) accessed at least one lecture. Among both registrants and starters, rates of accessing at least one lecture ranged from about one fourth to about two thirds. Across the 16 courses, the median rate of accessing the first lecture was 42% for registrants and 38% for starters for both the sequential and user-driven approaches.

With the high percentage of MOOC users who did not access any lecture, the retention rate offers another approach to identifying milestones. Table 2 shows that according to the sequential-based approach (and thus assuming that users access course materials in the order specified by the instructor), between 11% and 36% of registrants who accessed a lecture in the first module also accessed a lecture in the final module. For the user-driven approach, retention rates for registrants and starters in each course were 1 to 4 percentage points higher than for the sequential approach, ranging across courses from 13% to 39%.

Figure 1 illustrates the percentages of registrants in the 13 courses with relevant data who accessed lectures sequentially in each module or week of the course.[6] Figure 1 shows that both the percentage of registrants who accessed a lecture in the first module and the rate of decline in access across modules varied across courses.

Despite these variations, and regardless of the course's length, the patterns of decay were similar. Between the first and second modules, the decline in the share of registrants who accessed each subsequent lecture was steeper, as measured by both percentage point and percentage declines than between any two subsequent modules. Across courses, the median decline in the share of registrants who accessed a lecture in the first module and who then accessed a lecture in the second module was 23 percentage points. In contrast, the median decline in the share accessing a lecture in the second and third modules

Table 1
Summary of User Milestones and Outcomes in Selected Courses

	Median		Minimum		Maximum	
User Group	**Sequential**	**User Driven**	**Sequential**	**User Driven**	**Sequential**	**User Driven**
Registrants						
Total *n*	37,909	37,909	13,228	13,228	99,451	99,451
Total	100%	100%	100%	100%	100%	100%
Accessed any lecture		46%		27%		68%
Accessed first lecture	42%	42%	26%	26%	62%	62%
Accessed last lecture	7%	8%	5%	6%	17%	18%
Attempted any quiz		16%		2%		37%
Attempted first quiz	21%	21%	4%	4%	35%	35%
Attempted last quiz	6%	6%	2%	0%	13%	28%
Final grade 80% or above	3%	4%	1%	0%	10%	12%
Starters						
Total *n*	29,071	29,071	12,478	12,478	70,450	70,450
% of registrants		78%		53%		95%
Total	100%	100%	100%	100%	100%	100%
Accessed any lecture		43%		26%		64%
Accessed first lecture	38%	38%	25%	25%	61%	61%
Accessed last lecture	8%	8%	5%	5%	18%	19%
Attempted any quiz		17%		2%		42%
Attempted first quiz	22%	22%	4%	4%	42%	42%
Attempted last quiz	6%	9%	2%	1%	16%	17%
Final grade 80% or above	4%	5%	1%	1%	10%	12%

Note. Registrant is defined as anyone who registers between the date the course officially opens for registration and 2 months after the course end date. *Starter* is a subset of registrants and is defined as a user who registers no more than 1 week after the course start date. Percentage of registrants is the number of starters divided by the number of registrants. Minimum and maximum calculations represent the smallest and largest percentage values, respectively, for each individual category across all courses. Full details of aggregate and course-specific data are available from the lead author on request.

was 5 percentage points. The percentage of users who accessed subsequent lectures sequentially in the order specified by the instructor continued to decline but declined less sharply and then nearly flattened out by the fourth module.[7]

Sequential Versus User-Driven Quiz Attempts

Considerably smaller proportions of registrants and starters attempted quizzes than accessed lectures. Attempting a quiz seems to be an important milestone, as Table 1 shows that fewer than one in five registrants and starters attempted at least one quiz (16% and 17%, respectively), with the share of registrants attempting at least one quiz ranging across courses from 2% to 37%. Between 4% and 35% of registrants attempted the first quiz (median = 21%) according to both the sequential and user-driven approaches.

Figure 2 shows the percentage of registrants who attempted quizzes in the ordered, sequential pattern identified by the instructor. The findings show that the share of registrants who attempted a quiz declines considerably between the first and second course module, declined again but at a less-steep rate between the second and third course module, and then stabilized. The decline in quiz attempts was less steep than the decline in lecture access, and the plateau occurred a bit earlier (in the third rather than fourth module), likely because, across courses, quiz attempt rates were lower than lecture access rates. The median decline in the share of registrants attempting quizzes in the first and second course modules was 7 percentage points. The median decline between the second and third course modules was 3 percentage points.

Variations Based on Course Characteristics

Notwithstanding the limitations described in the Methods section, we explored variations in outcomes based on various course characteristics. These analyses did not reveal any strong patterns. As an example, the scatterplot in Figure 3 shows no discernable relationship between the percentage of registrants who received a final grade of 80% or higher and the number of course weeks/modules.

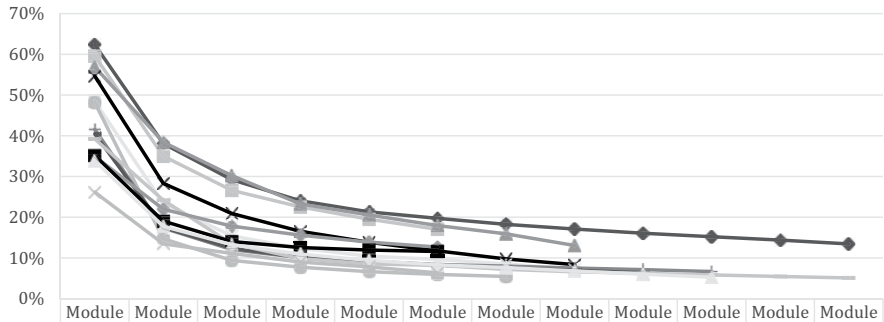
Discussion

Contemporary MOOC research is limited by the absence of consistent definitions of users and outcomes (DeBoer, Ho, Stump, & Breslow, 2014). This study advances knowledge by considering two definitions of MOOC users (registrants and starters), comparing two approaches to measuring student progress through a MOOC course (sequential versus user driven), and examining several measures of MOOC outcomes and milestones. We draw the following conclusions.

Table 2
Retention Rates of Registrants and Starters by Course (in percentages)

	Registrants		Starters	
Course	**Sequential**	**User Driven**	**Sequential**	**User Driven**
Cardiac Arrest, Resuscitation Science, and Hypothermia	36	39	37	39
Networked Life	33	37	35	39
Gamification	29	30	31	33
Rationing Medical Resources	24	26	25	26
Vaccines	23	24	26	27
Pay Attention: ADHD Through Lifespan	22	24	24	26
Greek and Roman Mythology	16	18	19	20
Principles of Microeconomics	16	17	17	18
Modern and Contemporary American Poetry	16	18	18	21
Health Policy and Affordable Care Act	15	18	18	20
Design: Creation of Artifacts in Society	15	17	16	18
Calculus: Single Variable	13	15	13	14
Listening to World Music	11	13	14	16
Median	16	18	19	21
Minimum	11	13	13	14
Maximum	36	39	37	39

Note. Retention rate is defined as the number of those who accessed a lecture in the last module divided by the number who accessed a lecture in the first module.

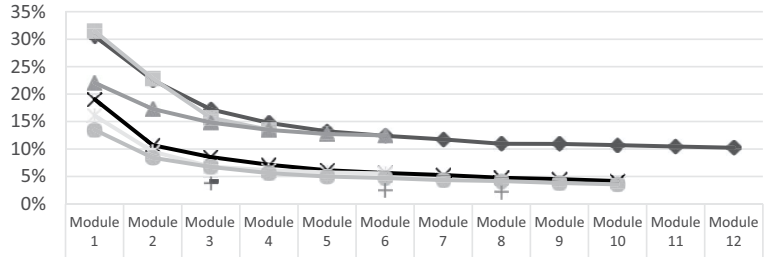


	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8	Module 9	Module 10	Module 11	Module 12
Pay attention!! ADHD through Lifespan	62%	38%	29%	24%	21%	20%	18%	17%	16%	15%	14%	13%
Gamification	60%	35%	27%	23%	19%	17%						
Vaccines	57%	38%	30%	23%	21%	18%	16%	13%				
Health Policy and Affordable Care Act	55%	28%	21%	17%	14%	12%	10%	8%				
Design: Creation of Artifacts in Society	48%	23%	15%	12%	10%	10%	8%	7%				
Listening World Music	48%	15%	9%	8%	7%	6%	5%					
Greek and Roman Mythology	42%	17%	12%	10%	9%	8%	8%	8%	7%	7%		
Principles of Microeconomics	40%	18%	12%	10%	9%	8%	8%	7%	6%			
Calculus: Single Variable	39%	24%	14%	10%	9%	8%	7%	7%	6%	6%	5%	5%
Cardiac Arrest	35%	22%	18%	16%	14%	13%						
Networked Life	35%	19%	14%	13%	12%	12%						
Modern and Contemporary American Poetry	34%	18%	13%	11%	9%	8%	8%	7%	6%	5%		
Rationing Medical Resources	26%	13%	11%	9%	8%	6%						

FIGURE 1. *"Sequential movement" of registrants: Percentage of registrants who accessed at least one lecture in each course module in the ordered sequence identified by the instructor (for the courses with relevant data available)*

First, the similarity of patterns for registrants and starters and using the sequential and user-driven approaches suggests that most users accessed course content in the sequential order identified by the instructor. Although a small percentage of users took advantage of the open nature of MOOCs to determine their own approach to accessing lectures, most users approached these first-generation courses in much the same way as a student would approach a course delivered in other formats. They tend to start at the beginning and work their way sequentially through a course.

Second, like other reports (e.g., Cusack, 2014; Ho et al., 2014; Koller et al., 2013), this study shows that only a very small



	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8	Module 9	Module 10	Module 11	Module 12
Pay attention!! ADHD through Lifespan	31%	23%	17%	15%	13%	12%	12%	11%	11%	11%	10%	10%
Gamification	31%	23%	16%	13%								
Cardiac Arrest	22%	17%	15%	13%	13%	12%						
Modern and Contemporary American Poetry	19%	11%	9%	7%	6%	6%	5%	5%	5%	4%		
Networked Life	16%	10%	7%	6%	6%	6%						
Greek and Roman Mythology	14%	8%	7%	6%	5%	5%	4%	4%	4%	4%		
Principles of Microeconomics			4%			2%		2%				
Listening World Music			4%									

FIGURE 2. *"Sequential movement" of registrants: Percentage of registrants who attempted a quiz in each course module in the sequence identified by the instructor (for the courses with relevant available data)*

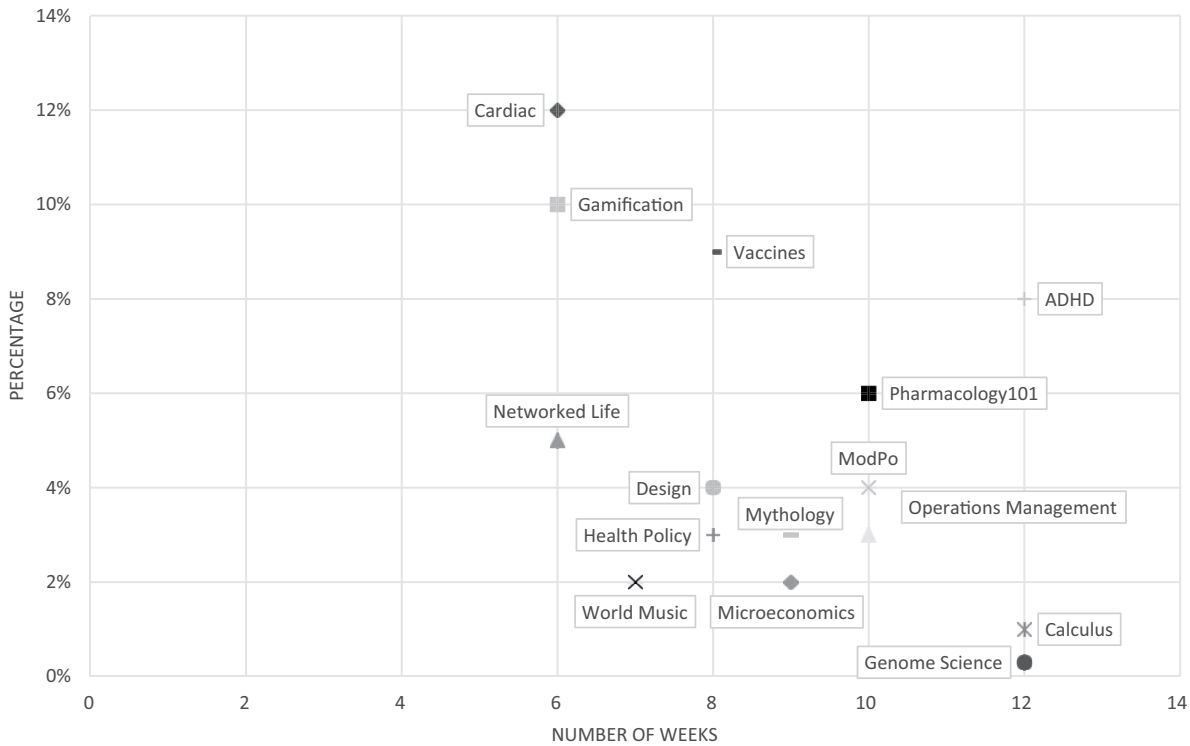


FIGURE 3. *Percentage of registrants who received final grade of 80% or higher by course length (calculated using user-driven approach to examining movement through the courses)*

share of users "completed" a MOOC. In no course did more than 12% of registrants receive a final grade of 80% or higher. Fewer than one in five registrants and starters in any course accessed a lecture in the last module. Retention rates were all below 40% even when calculated based on the number of users who accessed a lecture in the first module.

The low rates of "completion," accessing the first lecture, and attempting the first quiz, as well as the dramatic declines in lecture access and quiz attempts especially in the first few weeks of a course, may reflect curiosity, browsing, and lack of interest or motivation to complete, especially for these first-generation courses. Some (e.g., Koller et al., 2013) argue that completion rates are inappropriate measures of a MOOC as they fail to take into account users' intentions and enrollment goals. Community colleges in the United States offer a related argument, asserting that they enroll students with differing goals, including personal enrichment, short-term occupational training, transfer to a 4-year college, and completion of a certificate or associate's degree.[8]

Even though these low completion rates translate into a larger number of "completers" than for any conventional course (given the high number of registrants; Koller et al., 2013; Universities UK, 2013), low completion rates also suggest the potential utility of identifying milestones that predict completion. Ho et al. (2004) found that 50% of registrants in 17 Harvard and MIT MOOCs ceased activity in the 1st week, whereas 16% of the remaining participants ceased activity in the 2nd week. The share of registrants who ceased activity continued to decline in subsequent weeks of a course. This study finds considerable drop-off between registering for a course and accessing the 1st week of course content, suggesting that accessing the first lecture is one key milestone. Although accessing the first lecture was an important predictor of completion, retention rates even for those who access the first lecture were also low. Attempting the first quiz was a more important milestone than accessing the first lecture, as a smaller share of users did the latter than the former. Those who accessed a lecture in the 4th week were likely to access subsequent lectures, suggesting that participation to this point in time is another important predictor of course completion.

Implications for Future Data Collection, Documentation, and Research

Given the evolution of MOOC offerings and the changing characteristics of participants, the patterns established in this study are likely to be true only temporarily. In addition to testing whether the patterns found in this examination of first-generation courses hold for subsequent offerings, this study has other implications to inform future research. Most generally, recognizing the data limitations described in the Methods section, greater collaboration between MOOC researchers and instructors would reduce the challenges experienced in this retrospective study, facilitate identification of what "works" in a MOOC, and promote opportunities to intentionally test the effectiveness of particular course characteristics and instructional strategies.

Future research would benefit from the intentional collection of data that enable examination of relevant and meaningful dimensions of users' course experiences and outcomes. We had access to vast amounts of data on the 16 courses. But simply having "big data" is insufficient (DeBoer et al., 2014). Future research requires "better data." Before better data can be systematically and consistently collected, protocols for data collection, documentation, and interpretation are needed to document and define variables and encourage shared standards for collecting and coding data (Ho et al., 2014). Future research should build on the definitions offered in this study to refine and further develop appropriate and useful definitions and measures of users and outcomes.

Future research would benefit from a more complete set of outcome measures. Indicators available for this study fall considerably short of what the U.S. Department of Education (Ochoa, 2011) requires for academic engagement in distance and correspondence education courses.[9] Measures are needed to understand the benefits that may result from sporadic participation, as even those who access only portions of a course's content may have meaningful learning experiences (Haggard, 2013; Ho et al., 2014). Whereas completion is typically valued by institutions (Ashby, 2004), future research that considers measures of students' satisfaction with the learning experience and employers' satisfaction with skills of students may also produce fruitful insights.

Additional research is needed to understand how course design characteristics and pedagogical practices influence user outcomes. Some portion of the low completion rates in this study is likely attributable to the failure of these first-generation courses to adequately engage students in the course content (Baxter, 2012) or utilize effective instructional practices.[10] Future research should consider whether these outcomes reflect insufficient interaction with students, the absence of effective pedagogical approaches, and/or failure to motivate students to learn (Creelman & Reneland-Forsman, 2013; Simpson, 2013). The challenges of examining variations in outcomes based on course characteristics found in this study suggest the need for studies that systematically vary particular course characteristics and test the effects of these variations on user outcomes. Attempts to isolate the effects of course characteristics on outcomes must take into account the self-selection of students into particular courses (Ashby, 2004).

More research is needed to understand variations in outcomes for MOOCs that are targeted toward users with particular learning goals. The targeted audience for the 16 courses in this study was typically not specified in course syllabi, although most seemed geared toward promoting occupational training or personal enrichment. Future research should consider how user outcomes vary for MOOCs that are specifically targeted to such groups as individuals seeking personal enrichment, adult learners needing vocational or occupational training, currently enrolled college students seeking supplemental learning resources, secondary school students looking to improve academic readiness for college, and prospective college students considering whether college is right for them (Universities UK, 2013).

The President's Council of Advisors on Science and Technology (2013) asserts that although "many questions and challenges remain" about MOOCs, this innovation has the potential to increase access to high-quality higher education at low cost. MOOCs also offer the possibility of "transforming education at all levels" by adapting material to learners' needs and consequently maximizing learning outcomes. Additional research should build on the foundation offered here to determine whether MOOCs can achieve their promise in addressing the many challenges facing higher education.

NOTES

An earlier version of this paper was presented at the MOOC Research Initiative conference in Arlington, Texas, in December 2013. The authors are grateful for the helpful feedback that they received on this study from its advisory board members, Kevin Bell, Meg Benke, Wallace Boston, and Deke Kassabian. This research was supported in part by the MOOC Research Initiative funded by the Gates Foundation and administered by Athabasca University, the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305890015, the Quantitative Division of Penn's Graduate School of Education, and Penn's Alliance for Higher Education and Democracy

(AHEAD). The opinions expressed are those of the authors and do not represent those of the funders.

[1]Some second- and later-generation courses also incorporated "community TAs," defined by Coursera as volunteers who perform well in one iteration of a course and then assist in providing feedback to users in later versions of the course.

[2]Users could typically download lectures and submit quizzes and other assessments multiple times.

[3]Coursera also provided files with data on course announcements and peer grading. We did not use these data in this study.

[4]Course syllabi did not consistently identify the threshold for passing a course, and we were unable to find consensus in the literature about the appropriate grade threshold. For instance, employers typically require employees to maintain a certain grade point average to receive any available tuition benefits (Carmichael, 2008). The minimum required for full benefits ranges from C to A (Landes, 2012). We did not consider forum posts, as Coursera did not provide us with access to these data and because of challenges associated with interpreting these data (as discussed by Brinton et al., 2013).

[5]The data set does not permit identification of unique users; an unknown number registered for more than one of the study courses.

[6]We did not attempt to take into account differences in course length in the analyses because we had no expectation that users considered the number of course segments when making decisions about accessing course content.

[7]Because of space limitations, graphical depictions of the patterns for starters and the user-driven approach are not presented. The results mirror those in the included figures and are available on request.

[8]Some research shows a positive relationship between motivation and massive open online course (MOOC) completion. In an examination of engagement in three computer science MOOCs, Kizilcec, Piech, and Schneider (2013) found that the percentage of survey respondents who reported taking a course to improve employability was somewhat higher among completers than among those who sampled, audited, or disengaged. Creelman and Reneland-Forsman (2013) found that completion rates in online courses at Sweden's Linnaeus University were higher for courses that were required for a degree program than for optional courses. Their surveys suggest that respondents who registered for an online course but were not active had other priorities, including employment obligations, other course work, and family responsibilities (Creelman & Reneland-Forsman, 2013). Individuals who make a financial investment may be more motivated to complete. Koller, Ng, Chuong, and Chen (2013) report that completion rates are considerably higher for those in Coursera's "Signature Track" than for those who are not. The $49 Signature Track offers the opportunity to receive an "official, verifiable" electronic certificate that may be shared "with employers and educational institutions through a secure URL" but was available in only three of the courses examined in this study. Although MOOC users are now systematically offered the option to "join Signature Track," fewer than 1% of registrants in the 16 courses in this study selected Signature Track, too few to conduct meaningful examination of the relationship between this incentive and user outcomes.

[9]The U.S. Department of Education specifies that academic engagement includes, at a minimum

> submitting an academic assignment; taking an exam, an interactive tutorial, or computer-assisted instruction; attending a study group that was assigned by the institution; contributing to an academic online discussion; and initiating contact with a faculty member to ask a question about the academic subject studied in the course. Merely logging into the electronic classroom does not constitute academic engagement. (Ochoa, 2011, p. 7)

[10]Research on face-to-face instruction suggests that merely attending a class session does not constitute meaningful engagement. In one of the most commonly cited models of student engagement, Kuh, Kinzie, Bridges, and Hayek (2006) assume that college student success depends on engagement in five "educationally-effective practices": academic challenge, active and collaborative learning, student-faculty interaction, enriching educational experiences, and supportive campus environment. Research on online and distance education also suggests the importance of pedagogical practices to user outcomes (Brooke, 2003; Creelman & Reneland-Forsman, 2013; Simpson, 2013; Swan, 2003).

REFERENCES

Ashby, A. (2004). Monitoring student retention in the open university: Definition, measurement, interpretation and action. *Open Learning*, *19*(1), 63–78.

Baxter, J. (2012). Who am I and what keeps me going? Profiling the distance learning student in higher education. *International Review of Research in Open and Distance Learning*, *13*(4), 107–129.

Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2012). *Interactive learning online at public universities: Evidence from randomized trials*. New York, NY: Ithaka S+R.

Brinton, C. G., Lam, H., Chiang, M., Liu, Z., Jain, S., & Wong, F. (2013). *Learning about social learning in MOOCs: From statistical analysis to generative model*. Unpublished manuscript.

Brooke, C. G. (2003). Where do you want to learn tomorrow? The paradox of the virtual university. In B. E. Kolko (Ed.), *Virtual publics: Policy and community in the electronic age* (pp. 265–285). New York, NY: Columbia University Press.

Carmichael, K. (2008, June 4). Employer tuition assistance programs. *Fastweb*. Retrieved from http://www.fastweb.com/financial-aid/articles/709-employer-tuition-assistance-programs

Carnoy, M., Jarillo Rabling, B., Castano-Munoz, J., Duart Montoliu, J. M., & Sancho-Vinuesa, T. (2012). Who attends and completes virtual universities: The case of the open University of Catalonia (UOC). *Higher Education*, *63*, 53–82.

Council for Adult and Experiential Learning. (2007). *Enrollment decisions and persistence of on-line learners in NACTEL and EPCE: An exploratory study*. Chicago, IL: Author. Retrieved from from http://www.cael.org/pdfs/SloanReport

Creelman, A., & Reneland-Forsman, L. (2013). Completion rates: A false trail to measure quality? Let's call in the HEROEs instead. *European Journal of Open, Distance and E-Learning*. Retrieved from http://www.eurodl.org/?p=archives&year=2013&halfyear=2&article=583

Cusack, A. (2014, January 10). A handy cheatsheet on MOOCs. *edSurge*. https://www.edsurge.com/n/2014-01-10-a-handy-cheat-sheet-on-moocs

Daniel, J. (2012, September 24). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility [Web log post]. Retrieved from http://blog4222.blogspot.ca/2012/09/making-sense-of-moocs-musings-in-maze.html

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course": Reconceptualizing educational variables for massive open online courses. *Educational Researcher*, *43*, 74–84.

Dodd, T. (2014, October 6). Top MOOC provider edX no longer free for all. *Financial Review*. Retrieved from http://www.afr.com/p/national/education/top_mooc_provider_edx_no_longer_FooMSmV3LdSQHYGKND4LoI

Downes, S. (2012). *Connectivism and connective knowledge: Essays on meaning and learning networks*. Ottawa, ON: National Research Council Canada. Retrieved from http://www.downes.ca/files/Connective_Knowledge-19May2012.pdf

Grainger, B. (2013). *Massive open online course (MOOC) report*. London, UK: University of London International Programmes. Retrieved from http://www.londoninternational.ac.uk

Haggard, S. (2013). *The maturing of the MOOC: Literature review of massive open online courses and other forms of online distance learning*. London, UK: Department for Business, Innovation and Skills.

Hart, C. (2012). Factors associated with student persistence in an online program of study: A review of the literature. *Journal of Interactive Online Learning*, *11*(1), 19–42. Retrieved from http://www.ncolr.org/jiol/issues/pdf/11.1.2.pdf

Hills, F. W. (2010). *An exploratory study of the impact of institutional policies and practices of community and technical colleges in Texas on student persistence in online courses*. Available from ProQuest Dissertations and Theses.

Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *Harvard X and MITx: The first year of open online courses* (HarvardX and MITx Working Paper No. 1). Cambridge, MA: Harvard University/MIT.

Hossler, D., & Gallagher, K. S. (1987). Studying college choice: A three-phase model and the implications for policy-makers. *College and University*, *2*, 207–221.

Jaggars, S. S., Edgecombe, N., & Stacey, G. W. (2013). *What we know about online course outcomes: Research overview*. Community College Research Center, Columbia University, New York, NY. Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/What-We-Know-About-Online-Course-Outcomes.pdf

King, M. A. (1972). Primary and secondary indicators of education. In A. Shonfield & S. Shaw (Eds.), *Social indicators and social policy* (pp. 53–66). London, UK: Heinemann Educational Books.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). *Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses*. Paper presented at the Third International Conference on Learning Analytics and Knowledge, LAK'13, Leuven, Belgium.

Koller, D., Ng, A., Chuong, D., & Chen, Z. (2013, May/June). Retention and intention in massive open online courses. *Educause Review*, pp. 62–63.

Kuh, G., Kinzie, J., Bridges, B.K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature*. Washington, DC: National Postsecondary Education Cooperative.

Landes, L. (2012, July 20). Tuition reimbursement: A benefit for some employees and employers. *Forbes*. Retrieved from http://www.forbes.com/sites/moneybuilder/2012/07/20/tuition-reimbursement-a-benefit-for-some-employees-and-employers/

Lease, A. J., & Brown, T. A. (2009). Distance learning past, present and future. *International Journal of Instructional Media*, *36*, 415–426.

McGuire, R. (2014). *The best MOOC provider: A review of Coursera, Udacity, and edX*. Retrieved from http://www.skilledup.com/blog/the-best-mooc-provider-a-review-of-coursera-udacity-and-edx/

Moloney, J. F., & Oakley, B. (2010). Scaling online education: Increasing access to higher education. *Journal of Asynchronous Learning Networks*, *14*, 55–70.

MOOCs @ Edinburgh Group. (2013). *MOOCS @ Edinburgh 2013, Report #1*. Edinburgh, Scotland: University of Edinburgh. Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/6683

Moore, C., Shulock, N., & Offenstein, J. (2009). *Steps to success: Analyzing milestone achievement to improve community college student outcomes*. Sacramento, CA: Institute for Higher Education Leadership and Policy.

Noel-Levitz. (2011). *National online learners priorities report*. Coralville, IA: Author.

Ochoa, E. M. (2011). *Guidance to institutions and accrediting agencies regarding a credit hour as defined in the final regulations published on October 29, 2010* (GEN-11-06). Washington, DC: U.S. Department of Education, Office of Postsecondary Education.

Open University. (2014). *Teaching and learning*. Retrieved from http://www.open.ac.uk/about/main/strategy/teaching-learning-ou

Perna, L. W., & Thomas, S. L. (2008). *Theoretical perspectives on student success: Understanding the contributions of the disciplines* (ASHE-Higher Education Reader Report, Vol. 34, No 1). San Francisco, CA: Jossey-Bass.

Picciano, A. G. (2012). *Report on the evaluation of the Alfred P. Sloan Foundation's anytime, anyplace learning program*. New York, NY: Sloan Consortium.

President's Council of Advisors on Science and Technology. (2013). Letter to president Barak Obama on the development of massive open online courses. Washington, DC: Author. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_edit_dec-2013.pdf?dm_i=1ZJN,26IAK,E29F2Z,7VXSC,1

Rhoads, R. (in press). *MOOCs, high technology, and higher learning*. Baltimore, MD: Johns Hopkins University Press.

Rogers, E. (1962). *Diffusion of innovation* (5th ed.). Glencoe, IL: Free Press.

Simpson, O. (2013). Student retention in distance education: Are we failing our students. *Open-Learning: The Journal of Open Distance and E-Learning*, *28*(2), 105–119.

Stone, R. (1970). *Mathematical models of the economy and other essays*. London, UK: Chapman and Hall.

Swan, K. (2003). Learning effectiveness online: What the research tells us. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online education, practice and direction* (pp. 13–45). Needham, MA: Sloan Center for Online Education.

Universities UK. (2013). Massive open online courses: Higher education's digital moment? London, UK: Author. Retrieved from http://www.universitiesuk.ac.uk

AUTHORS

LAURA W. PERNA, PhD, is the James S. Riepe Professor and Executive Director, Alliance for Higher Education and Democracy (AHEAD), at the University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104; *lperna@gse.upenn.edu*. Her research interests focus on understanding how social structures, educational practices, and public policies promote and limit college access and success, particularly for individuals from lower-income families and racial/ethnic minority groups.

ALAN RUBY is a senior fellow at the Alliance for Higher Education and Democracy, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia PA 19130; *alanruby@gse.upenn.edu*. His research focuses on globalization and higher education and secondary school reform in newly independent nations.

ROBERT F. BORUCH, PhD, is the University Trustee Chair Professor in the Graduate School of Education and the Statistics Department of the Wharton School at the University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104; *robertb@gse.upenn.edu*. His current research focuses on randomized controlled trials in different disciplines, big data, research ethics, and failure analysis.

NICOLE WANG is a research fellow at the University of Pennsylvania, Graduate School of Education, 3700 Walnut Street, Philadelphia, PA 19104; *ruiw@gse.upenn.edu*. Her research focuses on MOOCs, quantitative analysis, evaluation, and big data.

JANIE SCULL is a doctoral candidate in quantitative methods at the University of Pennsylvania, Graduate School of Education, 3700 Walnut Street, Philadelphia, PA, 19104; *scull@gse.upenn.edu*. Her research primarily employs large-scale randomized control trials to examine differential student learning.

SEHER AHMAD is a doctoral student in higher education at University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104; *sehera@gse.upenn.edu*. Her research focuses on gender, education, stratification, and quantitative methods.

CHAD EVANS is a graduate student in the Department of Sociology at the University of Pennsylvania, 113 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104; *echad@sas.upenn.edu*. His research focuses on education, family and quantitative methodology.

Manuscript received May 8, 2014
Revision received October 28, 2014
Accepted November 11, 2014