

HarvardX and MITx: Four Years of Open Online Courses

Fall 2012-Summer 2016



Inside:

December 23, 2016

- 1) 4-year trends
- 2) Demographics
- 3) Certification rates
- 4) Reruns & modules
- 5) Course networks
- 6) Curricular areas
- 7) Stated intentions
- 8) Teaching teachers
- 9) Time to certification

This report is a collaboration between the Research Committee for the Vice Provost for Advances in Learning at Harvard University and the Office of Digital Learning at MIT.

HarvardX and MITx: Four Years of Open Online Courses

Isaac Chuang
MIT

Andrew Ho
Harvard University

December 23, 2016

In 2014 and 2015, a joint research team from Harvard University and MIT released summary reports¹ describing the first two years of Harvard and MIT open online courses launched on the nonprofit learning platform, edX. These reports set expectations for the demographics and behavior of course participants and established an analytic framework for understanding the then-nascent online learning context known as the Massive Open Online Course (MOOC).

This “Year 4 Report” extends these earlier findings to four complete years of HarvardX and MITx courses on edX, resulting in one of the largest surveys of MOOCs to date: 290 courses, 245 thousand certificates, 4.5 million participants, 28 million participant-hours, and 2.3 billion events logged online. We present our findings in a series of nine exhibits that address questions about the evolution of the MOOC movement from its birth in 2012, through its current adolescence.

[Exhibit 1](#): How has total MOOC participation and certification grown over time?

Steadily. Certification has slowed since free certification was discontinued in early 2016.

[Exhibit 2](#): Who takes a MOOC, and what do they do?

Participants in a MOOC “classroom” are heterogeneous in background and intention. There is no physical classroom like it on earth.

[Exhibit 3](#): What percentage of MOOC users earn certificates?

A typical course certifies 60% of paying users.

[Exhibit 4](#): If a MOOC is repeated, does its enrollment grow or shrink?

A typical MOOC re-run has been 25% smaller than its previous version.

¹ Ho, A. D., Chuang, I., Reich, J., Coleman, C., Whitehill, J., Northcutt, C., Williams, J. J., Hansen, J., Lopez, G., & Petersen, R. (2015). HarvardX and MITx: Two years of open online courses (HarvardX Working Paper No. 10). Retrieved from ssrn.com/abstract=2586847.

Ho, A. D., Reich, J., Nesterko, S., Seaton, D., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The First Year of Open Online Courses (HarvardX and MITx Working Paper No. 1). Retrieved from ssrn.com/abstract=2381263.

Exhibit 5: What are the hubs of the MOOC curricular network?

Computer Science courses are the largest (among, e.g., science, history, health, and other subjects) and route more participants to other disciplinary areas than they receive.

Exhibit 6: How do MOOC demographics and activity differ across curricular areas?

Computer Science and STEM courses are comprised of younger, less female, more international, and less college-educated participants than other courses.

Exhibit 7: How many MOOC participants state their intention to earn a certificate, and do they?

Slightly more than half (54%) of survey respondents report wishing to earn a certificate. A typical course certifies 30% of these respondents.

Exhibit 8: How many teachers take MOOCs?

32% of survey respondents report being or having been a teacher, and 19% of these respondents report teaching the topic of the MOOC.

Exhibit 9: How much online time does it take for a MOOC participant to earn a certificate?

A typical MOOC certificate earner spends 29 hours interacting with online courseware. One out of every hundred earns a certificate spending less than 23 minutes online.

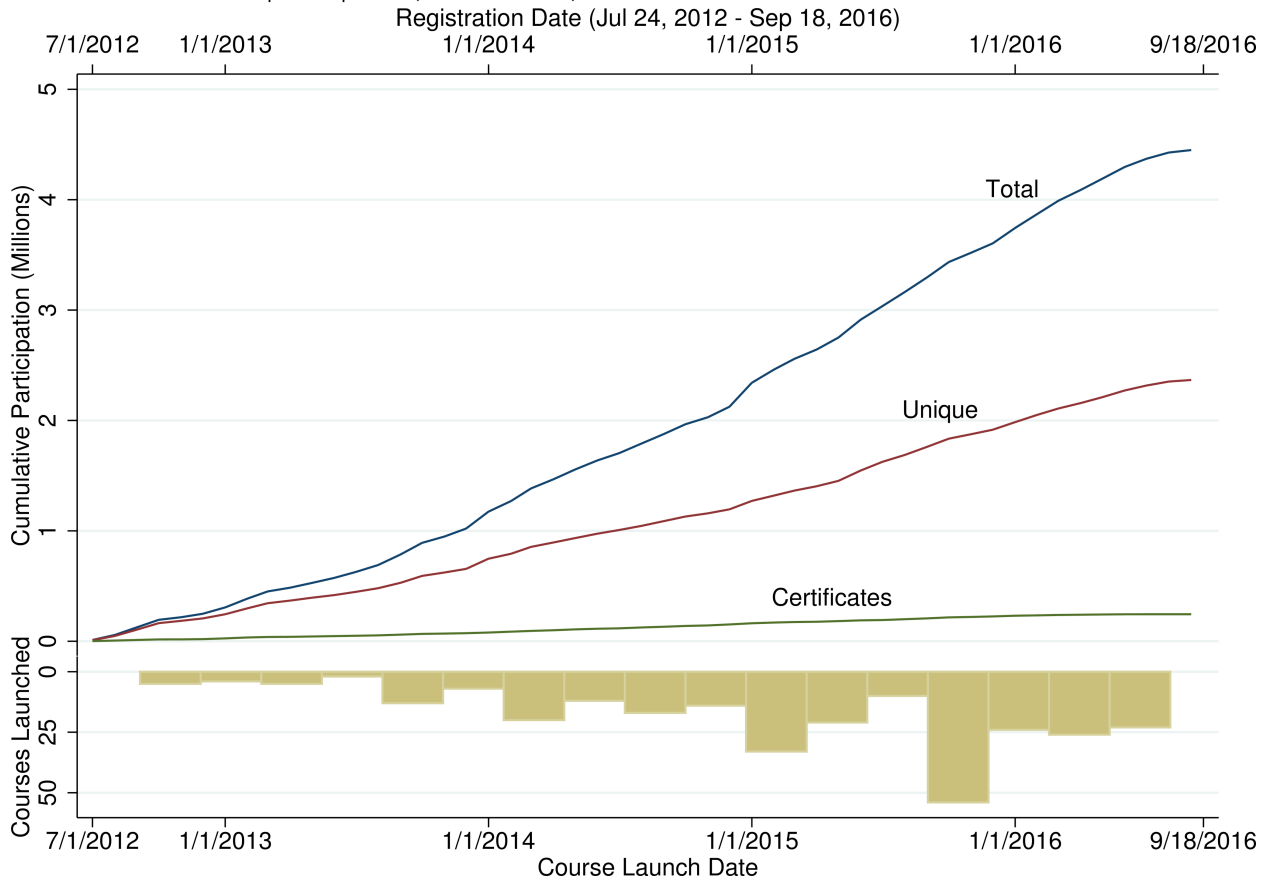
Data appendices enabling replication of many of these analyses, as well as many additional analyses, are available here: <http://year4.odl.mit.edu/appendix.html>

Exhibit 1: Cumulative trends in certification and participation

From the launch of HarvardX and MITx in the summer of 2012 through the fall of 2016, 245 thousand certificates have been issued, and 2.4 million unique users have participated in one or more HarvardX or MITx open online courses. Total course participation, counting participation in multiple courses, is 4.5 million.

- Cumulative unique and overall participation has grown steadily over this 4-year period.
- Around one third (31%) of unique participants have participated in multiple HarvardX and MITx courses.
- On average, 1,554 new, unique participants have enrolled per day, over 1,523 days.
- 245 thousand certificates have been earned by 159 thousand unique participants
- Cumulative certification has grown steadily until a policy change in early 2016 that largely discontinued free certification.

Exhibit 1: Cumulative participation, certification, and course launch dates over time.



- Small jumps are associated with the clumping of some course launch dates, usually around semester starting dates.
- Participants are those who have been active in the courseware. The figure excludes 3.89 million course registrants who never entered the courseware.

Exhibit 2: Course participation and demographics

Exhibit 2a: The size and makeup of a typical HarvardX and MITx open online course

Courses (including repeats and modules)	290
Certificates (free and paid)	244,705
Explorers (accessed >50% content)	739,260
Participants (accessed content)	4,449,034
Certificate earners per course	500
Explorers per course	1,517
Participants per course	7,902
Percentage of female participants	33%
Percentage of bachelor's degree holders	73%
Median age	29
Percentage from the United States	29%

Note: Per-course and demographic statistics are medians of course-level statistics.

- The typical HarvardX/MITx course or module certifies 500 participants.
- Around 1,500 will explore half or more of course content.
- Around 7,900 will access the course content.

Exhibit 2b: If a MOOC classroom held 100 students, who would they be?

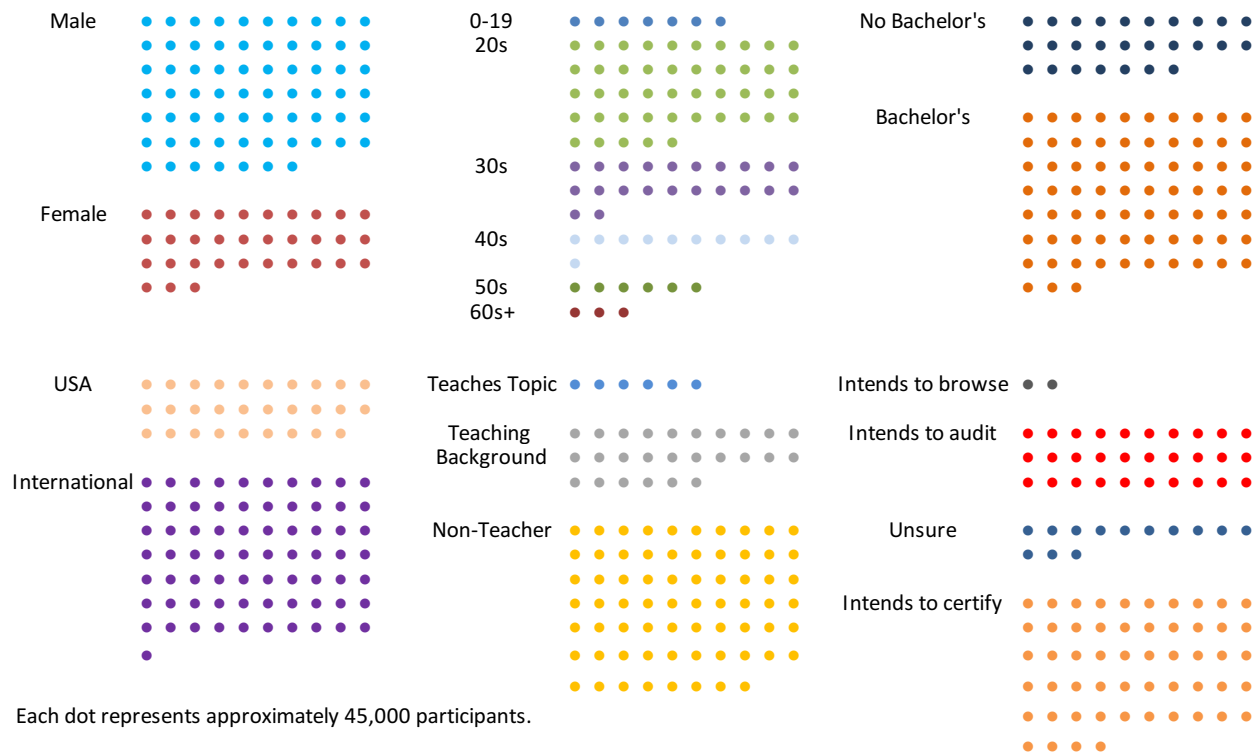
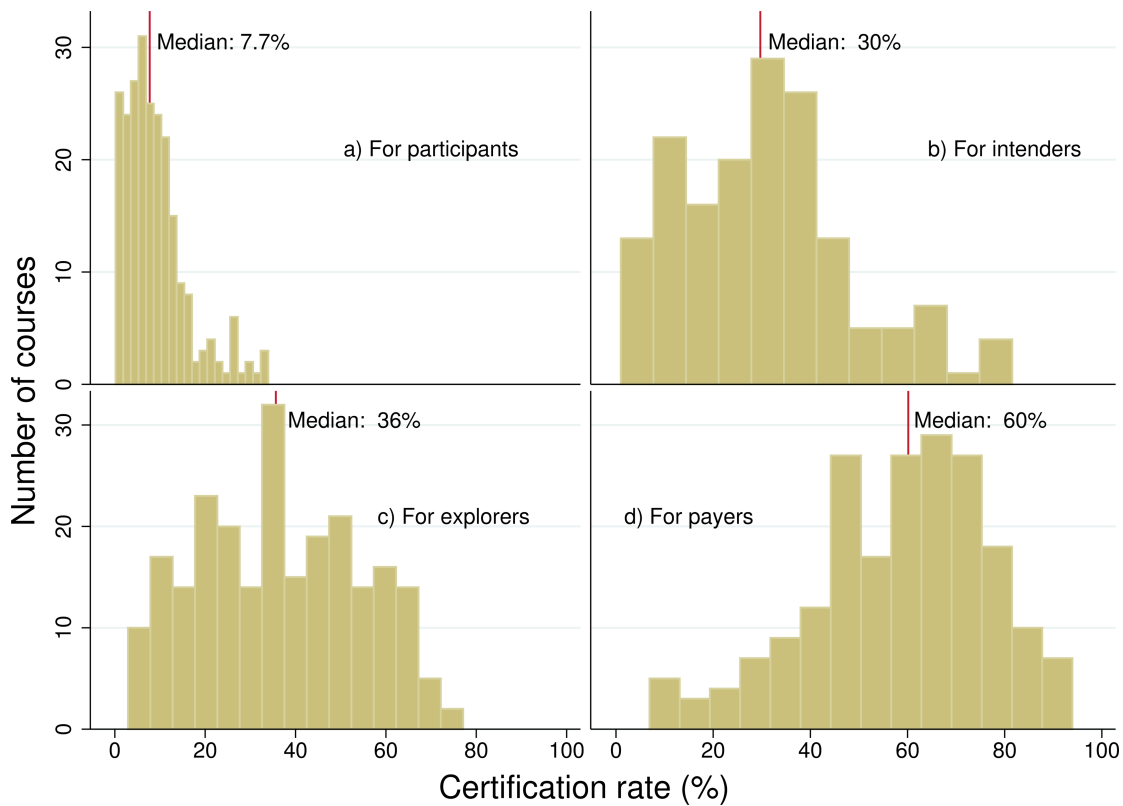


Exhibit 3: Certification rates

Certification rates are statistics that, properly contextualized, can describe and evaluate MOOCs. The naïve certification rate of 5.5% (245K certificates / 4,449K participants) is a useful reminder that not all participants earn certificates, but it is a poor evaluative metric for MOOCs. Not all MOOCs offer free certificates, and not all participants desire certificates. Exhibit 3a presents distributions of course certification rates that adjust for participation, certification, and course types.

Exhibit 3a: Course certification rates for participants, participants indicating intent to certify (“intenders”), explorers, and participants who pay for identity verification (“payers”).

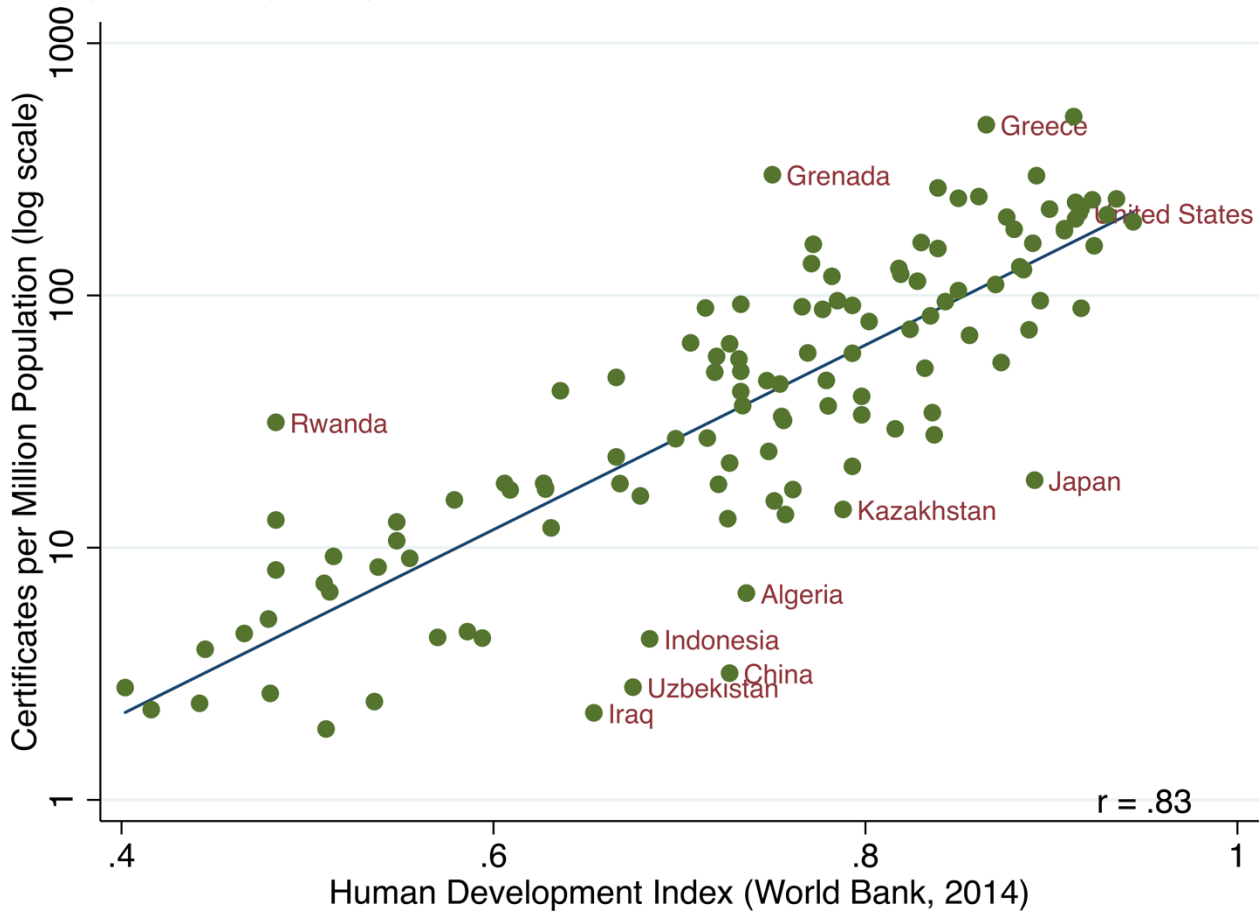


- The median certification rate is 7.7%, among 4.05 million participants in 236 HarvardX and MITx courses offering free certification. Certification rates range from 0.2% (Field Theory) to 34% (a Chinese History module). This rate differs from the naïve certification rate in part by excluding courses that do not offer free certificates.
- The median certification rate is 30% among 498 thousand participants who state that they plan on, “completing enough course activities to earn a certificate.” The survey was administered in 161 courses offering free certificates (see Exhibit 7). Rates range from 1% (CS50x) to 82% (a Chinese History module).
- The median certification rate is 36% among 662 thousand participants who “explore” a course (access at least half of course content) that offers free certification. Rates range from 3% (a course in Health Care Innovations) to 77% (the HarvardX course about ancient Greek heroes).

- The median certification rate is 60% among the 79 thousand participants who invest in a paid, “verified” certificate. 202 courses had at least 10 participants who took this option. Rates range from 7% (a Circuitry course) to 94% (a Chinese History module).

The per capita prevalence of HarvardX and MITx certification across countries is predicted well by the Human Development Index (HDI), a country-level composite of life expectancy, education, and income indicators. Controlling for population, HarvardX and MITx certificate earning remains disproportionately prevalent in countries with high HDI values where marketing, infrastructure, incentives, and supports are likely converging to enable participants to earn certificates. Nonetheless, there are outlying countries, like Rwanda, which has greater certification than expected by HDI, potentially explainable by targeted outreach there from a joint partnership between edX and Facebook.²

Exhibit 3b: The relationship between HarvardX and MITx certification prevalence and human development index, by country.



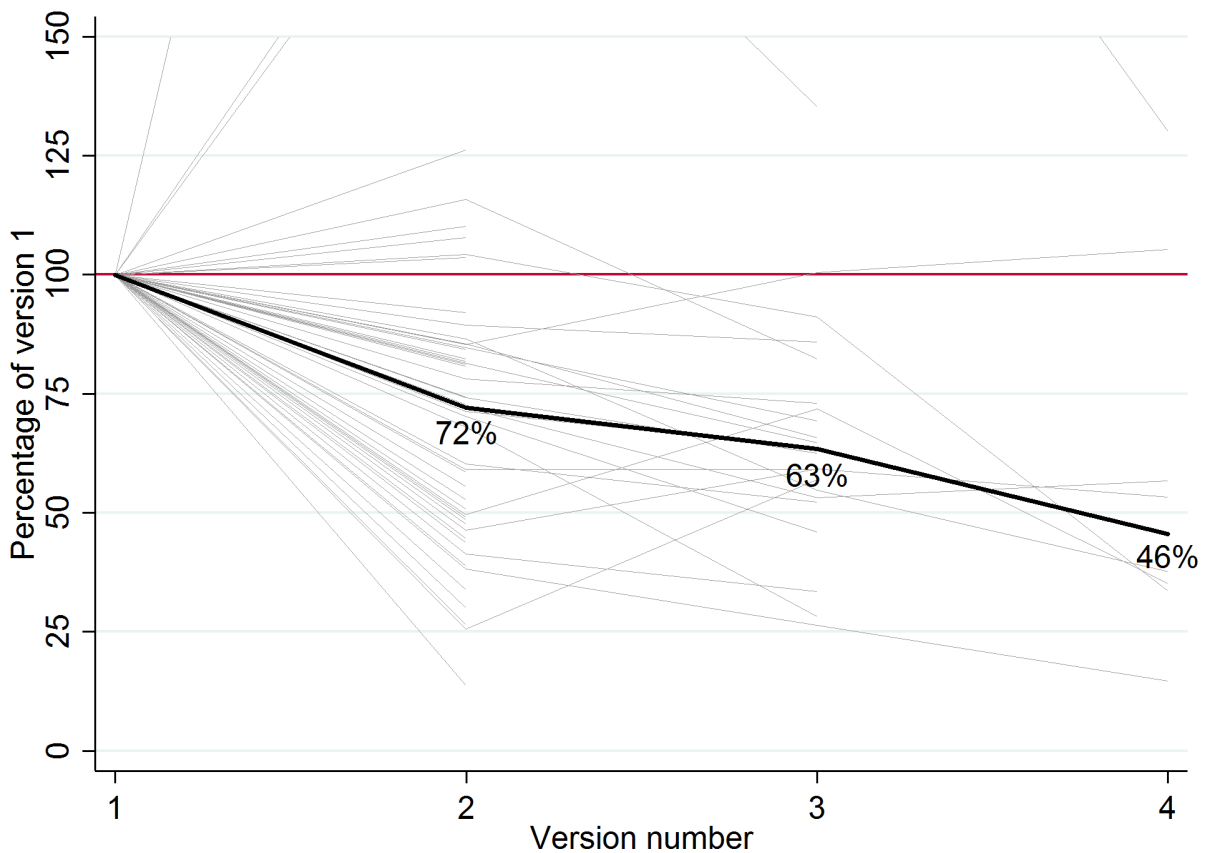
² <http://www.chronicle.com/blogs/wiredcampus/quickwire-edx-partners-with-facebook-to-offer-courses-in-rwanda/50693>

Exhibit 4: Enrollment trends for repeated and modularized courses

Repeated courses

Repeated courses have typical enrollment declines of around 25% from version to version. However, these declines vary widely by course. Some courses increase in enrollment and others decline by 50% or more. We have defined each repeat as its own course for the purposes of analysis. We call the course that is repeated (or modularized) the “parent course” (this is called a “subject” at MIT). See Exhibit 6 for additional trend data.

Exhibit 4a: Percentage of participants remaining in repeated runs of a “parent course,” with the median percentage shown in bold black.



- Over 4 years, among 96 “parent courses,” 52 have repeated at least once. Of these 52 parent courses, 29 have repeated once, 15 have repeated twice, and 8 have repeated 3 or more times.

Modularized courses

Many instructors and participants have opted for the flexibility of modularized courses that enable registration and certification in smaller fractions of course content. Among 96 parent courses, 23 have multiple modules. On average, participation drops off sharply by the second module, leveling off thereafter.

Exhibit 4b: Percentage of participants remaining in modules after the initial module, for the 23 parent courses with multiple module offerings.

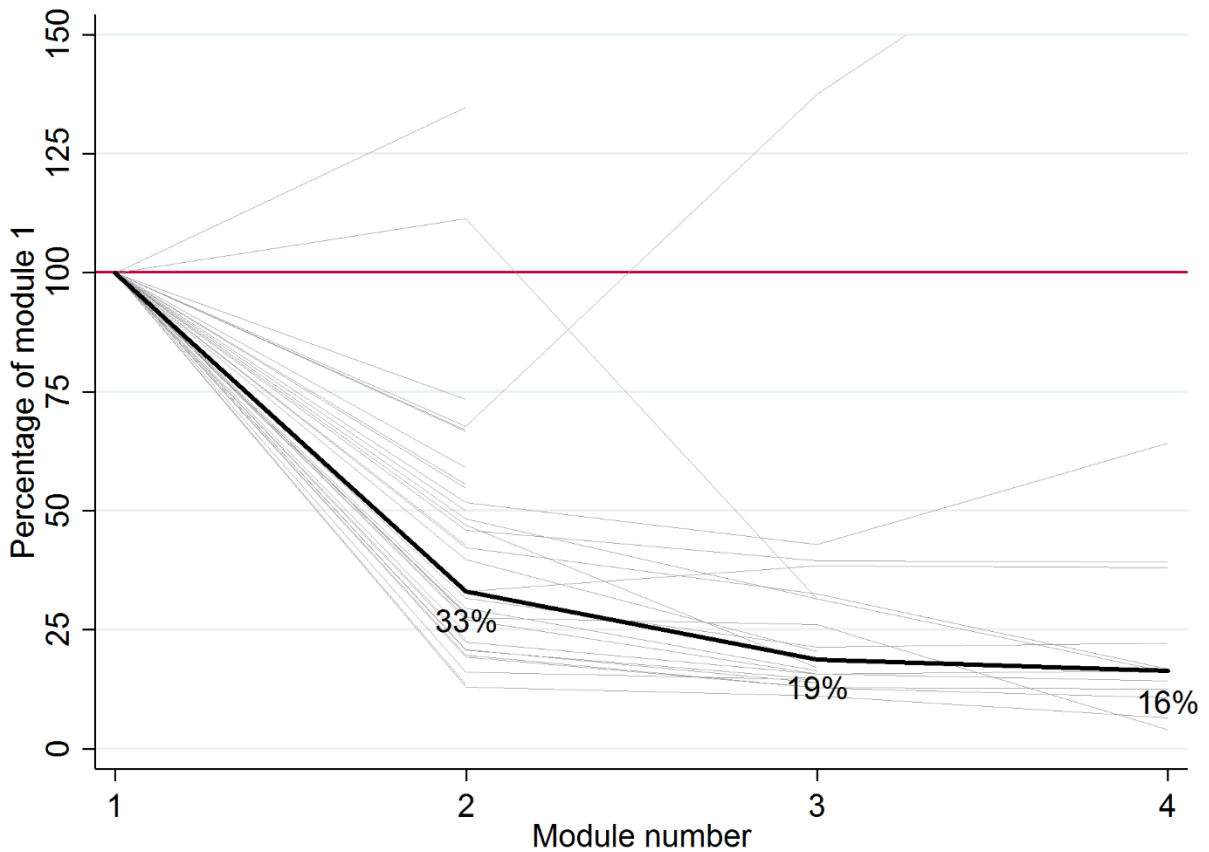
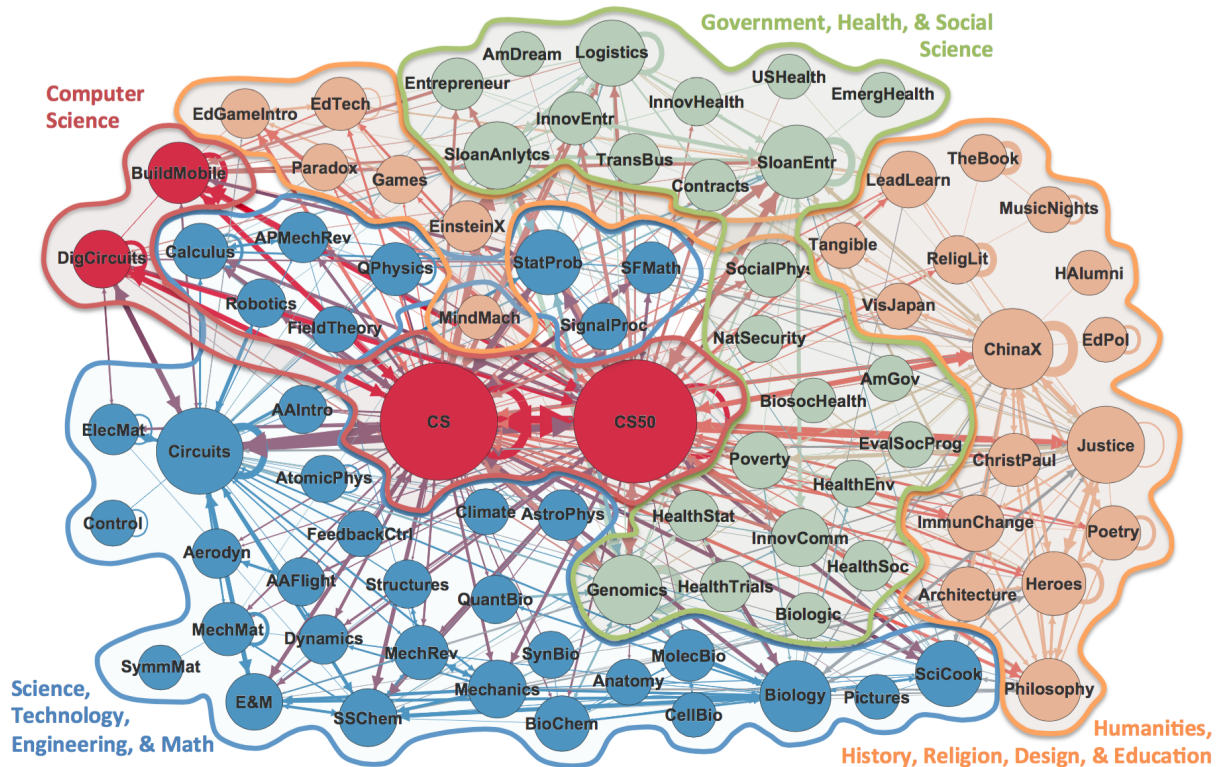


Exhibit 5: Flow of participants between courses

Participants who enroll in multiple courses over time can provide insight into the latent connections between courses and curricular areas.

Exhibit 5a. The HarvardX-MITx course network



Note: Larger nodes connect more participants to other courses. Central nodes connect to more courses. Proximal nodes have greater flow between them than distal nodes. Arrows indicate the direction of flow, with weights corresponding to the amount of flow.³ Estimated from 743K participants who participated in multiple courses.

- The HarvardX and MITx computer science (CS) courses are central to the network, with the most connections to other courses.
- As an illustration of the relevance of position, Circuits, at the left of the diagram, connects to many science, technology, engineering, and mathematics (STEM) courses but few humanities courses. Justice, at the right of the diagram, connects to many humanities courses but few STEM courses.
- Facilitators of curricular networks can use this information to build from existing connections, using introductory CS courses to route more deliberately to other curricular areas, or shore up

³ Nodes include all course versions and modules where applicable. Only paths with more than 50 participants are shown. We do not include any flow between courses when registrations occur within two hours of each other, to avoid counting casual participation in a number of courses over a short amount of time.

weak but substantive connections, like those between mathematics (at left) and philosophy (at right).

Exhibit 5b shows the course-to-course flow of participants across and within curricular areas. A typical CS course will send 772 participants to other CS courses, 514 to other STEM courses, 366 to other GHSS courses, and 239 to other HHRDE courses. CS courses both send and receive the most participants per course. A typical HHRDE course is just as likely to receive a participant from a CS course as another HHRDE course. In this way, CS courses continue to act as the active hub of MOOC networks.

Exhibit 5b. Average flow, in numbers of participants per course, within and between curricular areas.

	to CS	to STEM	to GHSS	to HHRDE	total sent
from CS	772	514	366	239	1891
from STEM	149	301	101	90	641
from GHSS	131	118	268	108	624
from HHRDE	60	81	85	233	460
total received	1111	1014	820	671	

Open online learning has enabled participants to enroll in courses across institutions without regard to conventional institutional boundaries. Exhibit 5c illustrates how participants have taken this opportunity across HarvardX and MITx courses. On an average, per-course basis, a HarvardX course will send 325 participants to another HarvardX course and even more, 334, to an MITx course. This is due in part to the large representation of participants from CS50x, which sends many participants to MIT courses subsequently. MITx courses also send substantial numbers of participants to HarvardX courses but almost twice as many, on average, to other MITx courses. This may reflect greater homogeneity of the MITx curricular area distribution, in comparison with HarvardX, which has many more courses in HHRDE and GHSS.

Exhibit 5c. Average flow, in numbers of participants per course, within and between institutions.

	to HarvardX	to MITx	total sent
from HarvardX	325	334	660
from MITx	251	492	743
total received	576	826	

Exhibit 6: Curricular and institutional breakdowns

The wide variation in participant behavior and demographics shown in Exhibit 2 is also evident across curricular areas. Exhibit 6a shows the number of courses and modules and their sizes and average demographic features.

Exhibit 6a. Course counts, sizes, and demographics, by curricular area.

	CS	STEM	HHRDE	GHSS
Courses	30	91	94	75
Certified	51,343	44,878	66,217	82,267
Explorers	171,502	163,235	163,729	240,794
Participants	1,527,300	1,081,995	822,026	1,017,713
Certified per course	1,523	333	443	651
Explorers per course	3,266	1,066	1,209	2,166
Participants per course	21,040	7,905	4,606	10,213
% Female	17%	16%	47%	35%
% Bachelor's	64%	63%	75%	84%
Median age	27	26	32	30
% USA	30%	25%	34%	27%

Note. CS=Computer Science; STEM=Science, Technology, Engineering, Mathematics; HHRDE=Humanities, History, Religion, Design, Education; GHSS=Government, Health, Social Sciences. Certification percentages and number certified per course are computed for the subset of courses that offered “honor certificates.”

- Participation and demographics by curricular area also show the relative draw of Computer Science among certifiers, explorers, and participants in open online courses.
- Government, Health, and Social Sciences (GHSS) courses are next largest on average.
- Humanities, History, Religion, Design, and Education (HHRDE) courses typically have slightly more certifiers and explorers than Science, Technology, Engineering, and Mathematics (STEM) courses, but slightly fewer initial participants. This may be because of the specialization of STEM courses at very high levels, compared to the more general interest nature of HHRDE courses.
- The percentage of female participants is quite low for CS and STEM courses.
- 84% of participants in the typical GHSS course have a Bachelor's degree. This is consistent with the education level of residential students taking these courses, many of which are offered by graduate schools.
- CS and STEM courses attract slightly younger participants.
- The typical STEM course has lower domestic participation than the typical CS and HHRDE course, though all curricular areas attract their majority of participants from abroad.

Exhibit 6b. Course counts, sizes, and demographics, across institutions and over time.

HarvardX	Total	Year 1	Year 2	Year 3	Year 4
Courses	129*	5	23	41	60
Certified	107,995	16,073	36,558	31,003	24,361
Explorers	366,161	38,572	83,289	138,632	105,668
Participants	2,100,298	289,540	441,202	827,131	542,425
Certified per course	502	2,743	1,737	453	330
Explorers per course	1,663	9,425	2,873	1,832	831
Participants per course	5,309	52,485	12,614	4,779	2,960
% Female	44%	43%	43%	42%	46%
% Bachelor's	76%	72%	79%	77%	75%
Median age	31	30	32	31	30
% USA	33%	30%	35%	32%	34%

MITx	Total	Year 1	Year 2	Year 3	Year 4
Courses	161**	11	27	51	72
Certified	136,710	27,252	34,056	46,737	28,665
Explorers	373,099	45,570	84,452	127,590	115,487
Participants	2,348,736	326,722	565,188	787,878	668,948
Certified per course	498	2,082	940	509	209
Explorers per course	1,398	3,543	2,357	1,567	951
Participants per course	10,837	28,535	21,063	12,208	5,645
% Female	20%	16%	19%	22%	19%
% Bachelor's	69%	61%	65%	72%	69%
Median age	27	27	28	28	26
% USA	25%	25%	27%	27%	23%

*HarvardX's 129 courses include 4 (3%) CS, 8 (6%) STEM, 80 (62%) HHRDE, and 37 (29%) GHSS.

**MITx's 161 courses include 26 (16%) CS, 83 (52%) STEM, 14 (9%) HHRDE, and 38 (24%) GHSS.

- HarvardX participants are older, have higher levels of education, and are more likely to be female than MITx participants on average. This is consistent with demographic differences in respective residential courses of the two parent institutions.

Exhibit 7: Intent

Certification rates that ignore participant intentions suffer from inappropriate comparisons (see Exhibit 3) to residential certification rates, where intention to certify is more consistent. This exhibit extends past analyses by Reich (2014) to report differential certification by stated intentions.⁴ Three quarters of HarvardX and MITx participants were sent a survey including the following prompt:

People register for HarvardX/MITx courses for different reasons. Which of the following best describes you?

- [Unsure] Have not decided whether I will complete any course activities.
- [Browse] Here to browse the materials, but not planning on completing any course activities (watching videos, reading text, answering problems, etc.).
- [Audit] Planning on completing some course activities, but not planning on earning a certificate.
- [Complete] Planning on completing enough course activities to earn a certificate.

Exhibit 7. Categories of response and nonresponse to a survey item about intention.

	Percentage	# Participants	% Explored	% Certified	% Female	Mean Age
Unsure	13%	127,908	18%	6%	34%	33
Browse	2%	22,213	23%	5%	32%	33
Audit	30%	299,972	24%	5%	34%	34
Complete	54%	537,714	31%	16%	30%	32
Responded	52%	987,807	27%	11%	32%	32
Did not respond	48%	896,599	15%	3%	34%	31
Surveyed	75%	1,884,406	21%	7%	33%	32
Not Surveyed	25%	628,292	15%	4%	34%	32

Note. Survey administered in 94 HarvardX and 101 MITx courses. Participants in an additional 99 courses were not given this item. In many cases, this is because participants were asked to fill in surveys only in the first of a sequence of modules.

- 54% of responding participants state an intention to earn a certificate (similar to previous years' reports). Of these 537,726 participants, 16% ultimately earn a certificate.
- These percentages are influenced considerably by CS50x, which accounts for 1/3 of those who intend to complete but has a certification rate around 1%.
- The typical course certification rate for those who intend to complete the course is 30% among the 161 courses that offered free certificates, as described in Exhibit 3.
- Certification rates for Browsers (5%) and Auditors (5%) are nonzero, even though participants in both groups state that they are not planning on completing activities or earning a certificate. This generally indicates that stated intentions are imperfect predictors of ultimate outcomes.

⁴ Reich, J. (2014). MOOC completion and retention in the context of student intent. *Educause Review*. Retrieved from <http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent>

Exhibit 8: Teachers

This exhibit extends past analyses by and Seaton et al. (2015) that showed high levels of participation by those with educational backgrounds, including teachers of the course content that is being taught.⁵ This finding was a useful indication of possible nontraditional pathways through which MOOCs may have a positive effect on educational outcomes, in this case by providing resources and learning opportunities to teachers. The standardized survey instrument includes two survey items relating to teaching, intended to capture the broadest sense of teacher and instructor identity:

- Are you currently or have you ever identified yourself as an instructor/teacher? Yes/No/Unsure
- Are you or have you ever taught material related to this course? Yes/No/Unsure

Exhibit 8. Categories of response and nonresponse to a survey item about teaching background.

	Percentage	# Participants	% Explored	% Certified	% Female	Mean Age
Teacher (is or has been)	32%	198,184	35%	16%	35%	38
Not a teacher (has never been)	68%	423,949	28%	12%	29%	30
Teaches this topic	19%	34,571	39%	17%	35%	39
Teaches another topic	81%	151,935	35%	16%	35%	37
Responded	46%	622,133	31%	13%	31%	33
Did not respond or unsure	54%	737,386	19%	4%	34%	31
Surveyed	54%	1,359,519	24%	8%	33%	32
Not Surveyed	46%	1,153,179	14%	4%	34%	32

Note. This survey item was administered in 83 HarvardX and 101 MITx courses.

- Teacher participation remains high, with 32% responding as being or having been a teacher.
- Of those self-described as being or having been a teacher, 19% teach the topic of the course.
- 16% of the self-described teachers certify; 12% of self-described non-teachers certify.

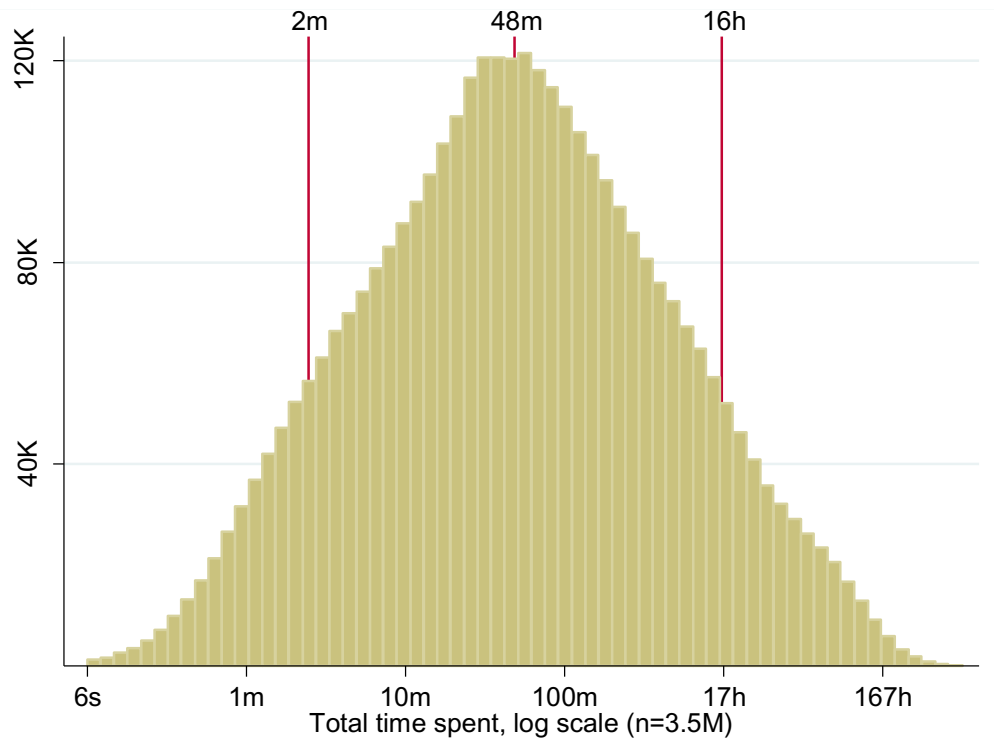
⁵ Seaton, D.T., Coleman, C., Davies, J., and Chuang, I. (2015) Enrollment in MITx MOOCs: Are we educating educators? *Educause Review*. Retrieved from <http://er.educause.edu/articles/2015/2/enrollment-in-mitx-moocs-are-we-educating-educators>

Exhibit 9: Comparing MOOC time online with on-campus credit hours

The “Carnegie unit” and the “credit hour” have become a nearly universal basis for measuring student progress in degree programs in US higher education. This unit, originally a metric of in-class “seat time,” can also describe the number of hours spent outside of lectures and discussions: completing assessments, attending labs, and preparing for class. Direct validation of this measure is difficult in a residential setting; attendance is recorded rarely, and students spend unmeasured time studying outside of the classroom, in libraries, dorm rooms, and coffee shops. Online courses offer new opportunities for estimating the time learners spend online toward certification.

Massive open online course components have direct corollaries in residential course components. Pre-recorded videos act as classroom lectures, online assessments act as periodic paper and problem set submissions, and archived discussion forums act as in-class or in-section discussions. Time spent completing these activities online can be estimated for participants from individual, server-logged events. Exhibit 9a shows estimated time spent for the 3.5 million participants with available server log data, counting any online activity as continuous as long as less than 30 minutes pass between server logged events.⁶ Under this definition of time spent online, Exhibit 9a shows that many participants browse courses for mere seconds while many others spend days.

Exhibit 9a. Distribution of total time spent online by participants in HarvardX and MITx MOOCs, on a logarithmic scale, with 10th, 50th, and 90th percentiles shown.

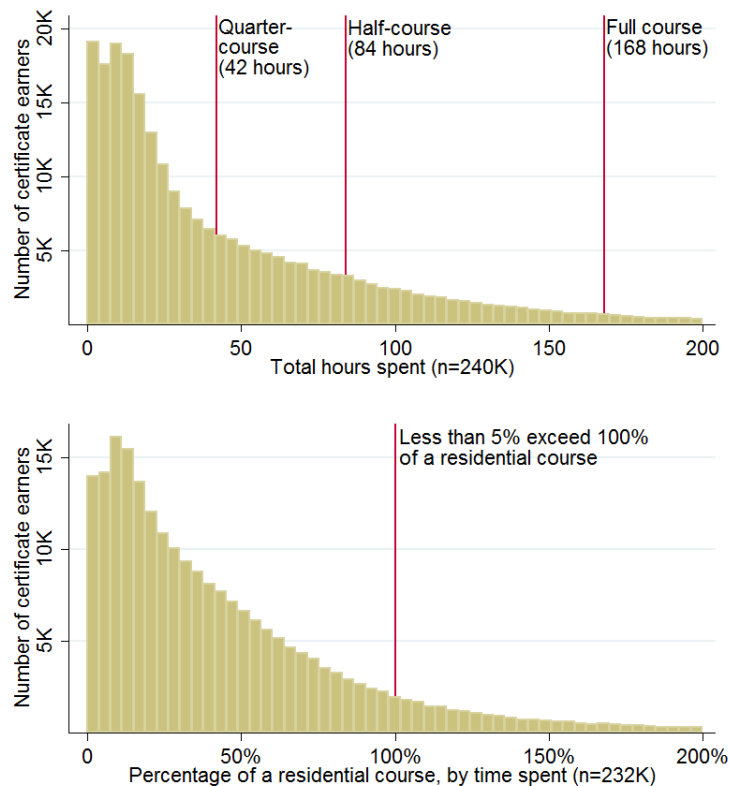


⁶ This is an industry-standard definition used in online analyses (Mullaney, 2014). Narrowing or widening the 30-minute window shifts the distribution but not enough to change key conclusions.

How much time do certificate earners, rather than overall participants, spend online? Exhibit 9b shows that most spend less than 50 hours. The 10th, 50th, and 90th percentiles are 5 hours, 29 hours, and 118 hours, respectively. One percent of certificate earners spend less than 23 minutes earning a certificate online.⁷ Extrapolating this to 4 courses per semester over 4 years, and a 4-year college degree could arguably be completed in 12 hours. Of course, these estimates of time spent online do not reflect time spent on offline activities. Actual times may be much higher if the course requires completion of offline reading, writing, or other tasks not directly instrumented in the MOOC courseware.

We compare MOOC time to residential course time by dividing certificate-earners' time spent online by the hours corresponding to the residential equivalent: 168 hours for a full course (12 hours per week over 14 weeks), 84 hours for a half-course, etc. The lower half of Exhibit 9b shows that 95% of certificate earners spend less time in a MOOC course than its residential equivalent, with a median of 34%. In-class seat time for a one-semester course generally accounts for 20%-35% of total time. Even under the conservative assumption that we cannot measure any time spent completing offline assessments, MOOC time online appears to be less than seat time for a substantial fraction of certified participants.

Exhibit 9b. Time spent by certificate earners in hours online (above) and as a proportion of the corresponding residential “course credit hours” (below).

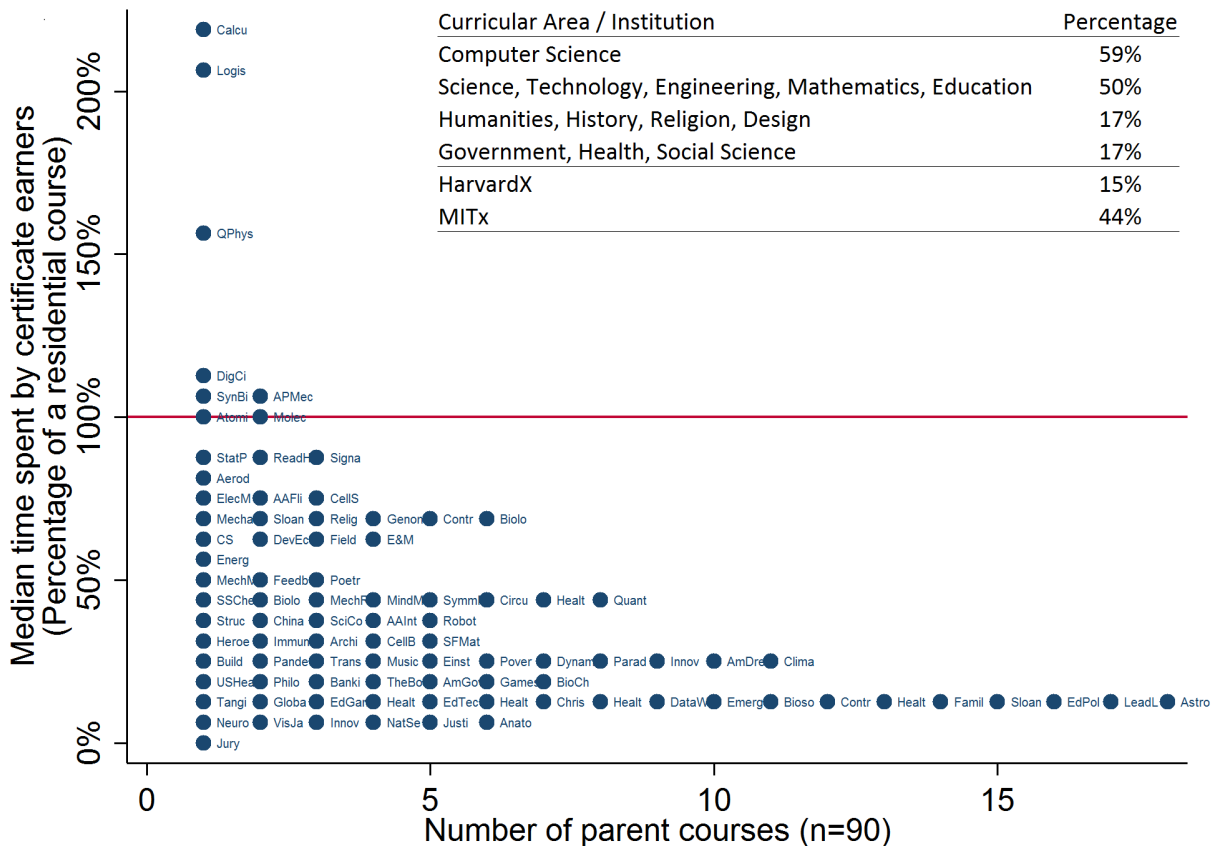


Note. Residential comparisons available for 272 courses (90 parent courses).

⁷ It is also possible that some individuals were gaming the system to earn certificates with little real effort, using strategies described by Northcutt, Chuang, and Ho (2016); however, these strategies only account for a small fraction of certificates earned. See <http://dx.doi.org/10.1016/j.compedu.2016.04.008>.

In terms of time spent online, what percentage of a residential course is a CS MOOC? A humanities MOOC? Exhibit 9c shows that the median participant in most courses spends a fraction of time online compared to time intended in residential courses, only 30% of the time of a typical course. This varies considerably by curricular area, with typical participants in a typical CS or STEM course spending around half the time of a residential course, and typical participants in a typical HHRDE or GHSS course spending around 1/6 of the time. The institution is a predictive factor, as well, with considerably less time online required by certificate earners in HarvardX courses than MITx courses.

Exhibit 9c. Median time online spent by certificate earners, by course, curricular area, and institution, as a fraction of residential course time.



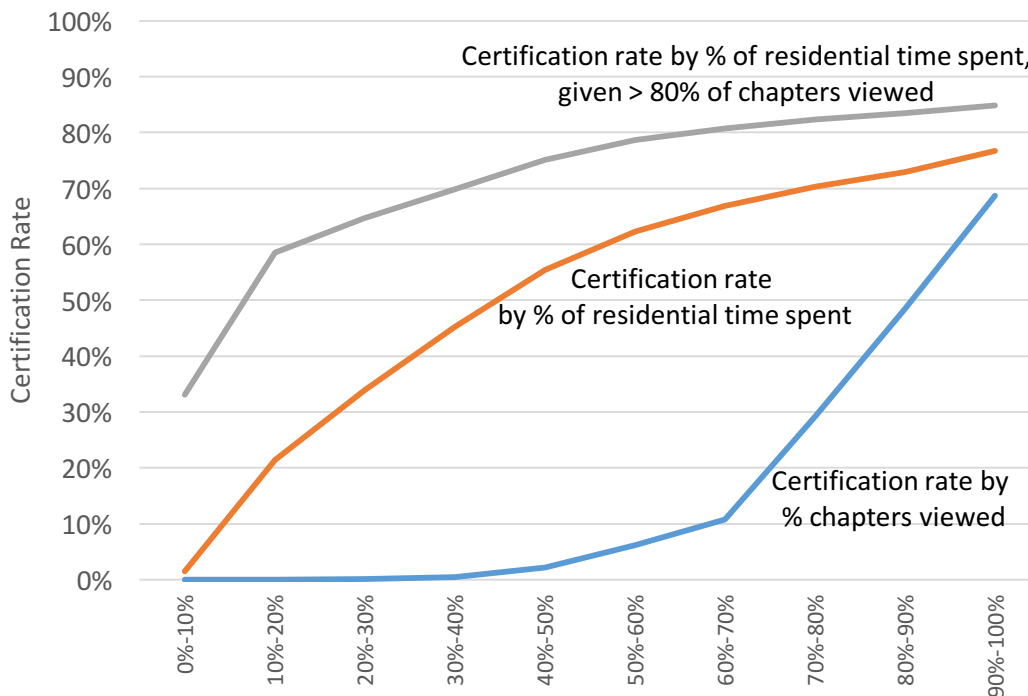
The time that participants spend online is also a powerful predictor of certification, like many other activity statistics in MOOCs. The central line in Exhibit 9d shows that almost no one who spends less than 10% of residential time online earns a certificate, but 21% of those who spend 10%-20% of residential time online earn a certificate. Those who spend half or more of residential time online have certification rates from 62% to 77% as time online increases.

For reference, the lower line in Exhibit 9d shows that accessing the breadth of course content does not predict high likelihoods of certification until participants access 80% or more of course chapters. This is unsurprising given the design of most courses to require assessments from throughout the course

content in order to earn certification. The top line in Exhibit 9d restricts the time analysis to participants who access more than 80% of the course chapters. For these participants, we observe certification rates near 60% even when participants spend only 10%-20% of residential course time online. For participants who spend more than 50% of residential course time online, certification rates climb above 80%.

Exhibit 9d highlights that participants whose observable behavior is similar to that intended for residential students, spending requisite time and surveying requisite content, certification rates are similar to course completion rates in residential contexts. Further, certification rates around 60% are observable among participants who spend much less residential course time online and still survey the breadth of course content.

Exhibit 9d. Certification rates by percentages of residential time and percentages of chapters viewed.



Note. Due to large sample sizes, standard errors are never more than half a percentage point. Analysis restricted to courses offering honor certificates (3.01 million participants).