

Bringing student backgrounds online: MOOC user demographics, site usage, and online learning

Jennifer DeBoer¹
jdeboer@mit.edu

Andrew Ho²
Andrew_Ho@gse.harvard.edu

¹Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Glenda S. Stump¹
gsstump@mit.edu

David E. Pritchard¹
dpritch@mit.edu

Daniel Seaton¹
dseaton@mit.edu

Lori Breslow¹
lrb@mit.edu

²Harvard Graduate School of Education
455 Gutman Library, 6 Appian Way
Cambridge, MA 02138

ABSTRACT

MOOCs gather a rich array of click-stream information from students who interact with the platform. However, without student background information, inferences do not take advantage of a deeper understanding of students' prior experiences, motivation, and home environment. In this poster, we investigate the predictive power of student background factors as well as student experiences with learning materials provided in the first MITx course, "Circuits and Electronics." We focus on a group of survey completers who were given background questions, and we use multiple regression methods to investigate the relationship between achievement, online resource use, and student background. Online course providers may be able to better tailor online experiences to students when they know how background characteristics mediate the online experience.

Keywords

MOOCs, social capital, regression, online learning

1. RATIONALE

Massive Open Online Courses (MOOCs) enroll thousands of students through numerous learning platforms. edX, which offered its first class "Circuits and Electronics" (6.002x), in the spring of 2012, drew students from nearly every country in the world [5]. The courses' free, open-enrollment structure appeals to a diverse set of students who can use learning resources as they please. But while on-campus instructors have the opportunity to learn about their students—their experience growing up, their educational background, etc.—as they interact with them in the classroom, MOOC instructors do not have that luxury.

MOOC students, however, are unique individuals. They are motivated by different incentives to sign up for the course, they come from different home environments, and they speak different languages. As part of a study on 6.002x, we gathered additional background information on students in the first edX class in order to understand the simultaneous impact of time spent on different course components and a student's background characteristics.

2. RESEARCH QUESTION

To provide a more complete picture of the student factors that relate to achievement, we ask, "What student factors predict higher achievement, all else equal?" In other work, we have explored the impact of resource use (e.g., watching videos, reading the textbook) on achievement in our analyses, but here we investigate whether the inclusion of demographic background factors as covariates changes the relationship between resource use and achievement.

3. CONCEPTUAL FRAMEWORK

We apply a social capital lens to our work. While the inputs of formal schooling are important, education researchers have also noted the importance of the acculturation and social preparation to which students are differentially exposed prior to entering school

[see, for example, 3, 6]. Once students enroll, the norms and behaviors they understand as beneficial may be differentially rewarded within the school system [1, 8]. The tools that serve more privileged groups of students in traditional settings (e.g., linguistic capital, knowledge of cultural references, highly educated role models) may also be relevant for online learning.

Distance learning classes have largely been characterized by learners looking for flexibility, cost-savings, and a familiar electronic platform [9, 5]. A larger portion of these students has been female, and many of them have been "non-traditional" students. While this suggests the online context may better support students who are underserved in traditional STEM classrooms, initial results from more recent online learning programs suggest there are no differential gains for underserved subgroups [2]. In fact, another study notes that the "achievement gap" between traditionally higher- and lower-performing students may actually be widened due to an online course experience [10].

4. DATA

Data come from the students in 6.002x who completed the exit survey. While the survey was announced specifically for course completers, the link to the survey was open on the website. We find ~800 completers did not receive a certificate in the course. It is important to note these data were gathered using matrix sampling to mitigate non-response due to survey fatigue. We impute missing responses using chained equations [7]. Missingness ranges from 59% to 85% of the over 7,000 students who completed the survey.

Figure 1 illustrates the spread of total points awarded to survey completers using a "partial credit" model. In this model, the number of points students received for a right answer was dependent upon the number of attempts they made. (Students were allowed unlimited attempts to answer homework and lab questions, but only three attempts on the midterm or final.) The bimodal distribution illustrates the "two-population" nature of our sampling frame: some students who earned a certificate, and others who followed the course but did not earn a certificate.

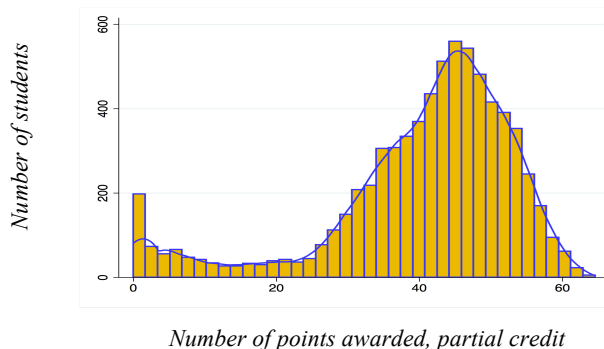


Figure 1. Survey completer numbers across performance levels

5. RESULTS

For the subset of students completing surveys, once we control for key student background information, we immediately find the impact of certain resources on students' total score to be diminished. In addition, when we add controls such as score on the first homework (proxy for prior ability), and when we control for the students' country, we find that we remove more bias. This may result, for example, from correlation between initial score and subsequent study strategies such as referencing previous homeworks or viewing relevant questions on the discussion forum.

Table 1. Additive models predicting partial credit score, OLS¹

Covariate	"Naïve" estimate	Control for first HW	Controls for country (not listed)
Homework	4.41 (0.28) ***	3.70 (0.26) ***	3.71 (0.27) ***
Labs	0.61 (0.34)	0.45 (0.32)	0.52 (0.32)
Lecture problems	0.26 (0.10) **	0.09 (0.09)	0.10 (0.09)
Lecture videos	0.38 (0.13) **	0.17 (0.13)	0.09 (0.13)
Tutorials	0.11 (0.06)	0.09 (0.06)	0.07 (0.06)
Book	-0.26 (0.07) ***	-0.23 (0.07) **	-0.24 (0.07) ***
Wiki	-0.64 (0.07) ***	-0.60 (0.07) ***	-0.58 (0.07) ***
Discussion board	0.30 (0.12) **	0.39 (0.11) ***	0.33 (0.11) **
Female	-1.11(1.80)	-1.20 (1.88)	-1.12 (1.88)
Parent engineer	2.16 (0.81) **	1.86 (0.79) **	1.91 (0.79) **
Worked with other offline	2.05 (0.66) **	1.96 (0.65) **	2.11 (0.71) **
Teach EE	-0.01(0.56)	0.08 (0.58)	0.26 (0.58)
Took diff. equations	4.72 (0.56) ***	4.44 (0.48) ***	4.56 (0.55) ***
First HW		0.57 (0.03) ***	0.55 (0.03) ***

5.1 Course resources

For survey completers, time spent on homework was a consistent significant predictor of a higher overall score. Even controlling for initial performance, spending more time on the homework was related to gains of approximately 1/3 of a standard deviation on the total points for the class (a small-to-medium effect size in educational research). More time spent on the discussion board was also related to a higher score. However, more time spent on the book or the course wiki was related to lower achievement.

5.2 Initial score

We construct a control covariate of "firstpoints," a rough proxy for initial proficiency with material relevant to the course. In the second and third models given above (additive), the inclusion of this control alters the significance of the covariates for time spent

¹Note: resource use covariates given in log-seconds. As noted above, partial credit is given for multiple attempts at a question, though results are consistent when full points were awarded regardless of the number of attempts. This was the policy used by the instructors in determining grades.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

on lecture problems and time spent on lecture videos. This may indicate that students who come into the course with different abilities use these resources in different ways. However, including this control does not change which demographics are significant.

5.3 Demographics

The impact of key demographic background factors is consistent across models, including the fully specified model with all covariates, which allows for a fixed-effect for the student's country of access. Individual factors such as gender and whether the student teaches electrical engineering are not related to achievement. On the other hand, some background factors are strongly related to performance. Specifically, having taken differential equations predicted a higher score, even controlling for the first assignment. Similarly, students who reported offline collaboration also scored higher. This might reflect the same positive role of collaboration as participation in the discussion forum.

6. DISCUSSION AND IMPLICATIONS

The inclusion of demographic variables for MOOC users adds significant, practically important covariates to predictions of achievement based on individual information. While online course creators may already have a wealth of student data from click-stream information, solely predicting performance based on observed behaviors misses important explanatory factors and a deeper understanding of why students may behave in different ways or experience differential utility of online resources. As MOOC offerings grow, course designers may further study how to tailor the online experience and support the diverse backgrounds of a world of students.

7. ACKNOWLEDGMENTS

Funding was provided by NSF Grant, DRL-1258448. We also acknowledge the support of MIT's RELATE group and edX.

8. REFERENCES

- [1] Anyon, J. 2008. *Social Class and School Knowledge*, in *The Way Class Works*, L. Weis (ed). Routledge, chap. 13.
- [2] Bowen, W.G., Chingos, M.M., Lack, K.A., & Nygren, T.I. 2012. Interactive Learning Online at Public Universities: Evidence from Randomized Trials, ITHAKA, May 22, 2012.
- [3] Bourdieu, P. 1977. *Cultural Reproduction and Social Reproduction*, in *Power and Ideology*, J. Karabel & A.H. Halsey (eds), Oxford Press, Chap. 29.
- [4] Dziuban, C., Moskal, P., Brophy, J., Shea, P. 2007. Student satisfaction with asynchronous learning, *Journal of Asynchronous Learning Networks*, 11(1), 87-95.
- [5] edX. 2012. Frequently asked Questions. *From: [edx.org/faq]*.
- [6] Lareau, A. 2003. *Unequal Childhoods: Class, Race, and Family Life*. University of California Press.
- [7] Little, R.J.A. & Rubin, D.B. 2003. Statistical Analysis with Missing Data. *Technometrics*, 45 (4), pp. 364-365.
- [8] Meyer, J. 1977. The Effects of Education as an Institution, *American Journal of Sociology*, 83(1): 55-76.
- [9] Park, J.-H., & Choi, H. J. 2009. Factors influencing adult learners' decision to drop out or persist in online learning. *Educational Technology & Society*, 12(4), 207-217.
- [10] Xu, D. & Jaggars, S.S. 2013. *Adaptability to Online Learning: Differences Across Types of Students and Academic Subject Areas*. CCRC, Working paper No 5.