Diversity in MOOC Students' Backgrounds and Behaviors in Relationship to Performance in 6.002x

Jennifer DeBoer
MIT Teaching and Learning Laboratory
jdeboer@mit.edu

Glenda S. Stump
MIT Teaching and Learning Laboratory
gsstump@mit.edu

Daniel Seaton
MIT Office of Digital Learning
dseaton@mit.edu

Lori Breslow
MIT Teaching and Learning Laboratory
lrb@mit.edu

Abstract

The first edX course had over 150,000 students enrolled, which included registrants from nearly every country in the world, bringing with them massive international diversity. These students were also diverse on a number of background characteristics. To augment the behavioral and geographical location data available from edX clickstream data, we gathered detailed individual background data for a subsample of students who completed an exit survey. Furthermore, we show that student performance varies significantly with some of these background characteristics. Our descriptive work highlights the important challenges that such a diverse classroom poses for instructors, course designers, and education researchers.

1. Introduction

The new global wave of large virtual courses offered for free has attracted an incredibly diverse population of students. In this paper, we apply a descriptive lens to the first massive open online course offered by MITx, "6.002x: Circuits and Electronics." The doors to this class, traditionally taken by computer science and electrical engineering sophomores at MIT, were thrown open wide to the world. Participants in 6.002x included advanced engineers already practicing in the field, high school students in Mongolia, and casually interested learners in nearly every country. We demonstrate there is a high degree of variability in all measurable dimensions for the participating students. This variability poses challenges and opportunities for researchers, instructors, and course designers.

2. Research Questions

In this study, we ask, broadly, "What variability do we observe in the background characteristics of the students and in their use of the 6.002x site?" We focus on three important areas of the student experience in this paper. More specifically, we ask,

a. What is the variability in location and behavior surrounding site access and site use?
b. For students who completed the exit survey, what is their prior exposure to the content, and what is their familiarity with teaching the content?

c. For students who completed the exit survey, what are the reasons they cite for having taken part in the course?

3. Findings

We organize the findings of our study as follows. First, we describe the variability we observed in the location from which the users access the site. We then focus on the explanatory power of more detailed student background information from students who completed the exit survey. The exit survey includes prior educational experiences as well as relevant content experience and stated motivation for enrolling in the course. Most survey completers were also certificate earners, though numerous students who were still active on the site at the end of the course in June also noticed the announcement about the survey and responded to it.

Although the sampling frame for this study comprises students who were given the exit survey, it is important to note that the survey was administered using matrix sampling. In other words, every student was given a random selection of questions from the survey, and, thus, some students were not given the opportunity to answer some questions. In figures below, "NA" denotes the students who did not receive the associated question on the exit survey. We therefore conduct one-way analysis of variance tests on the sub-sample of students who were administered four important questions related to offline collaboration, educational attainment, experience teaching this content, and reason for enrolling in the class. We illustrate important variation in performance by these four key background variables. Previous research on residential education indicates that these constructs are important in explaining variation in performance (e.g., benefits of collaboration: [1, 2]) and may also be significant in this new virtual learning environment.

3.1. Variability in location

First, we show that there is a large degree of geographic variability, indicating participation from students around the world. After determining the students' points of access via IP addresses, we found that students logged on to the site from nearly every country in the world. However, the level of participation was highly skewed, with only twelve countries individually accounting for greater than 2% of all participating students each. Most countries had less than one hundred participants. Furthermore, a significant number of students accessed the site from multiple locations. Our data showed that a number of 6.002x participants were highly mobile, and logged on from multiple countries. Whereas many online education programs have been geared towards local populations, MOOCs are a global opportunity for a globalized audience.

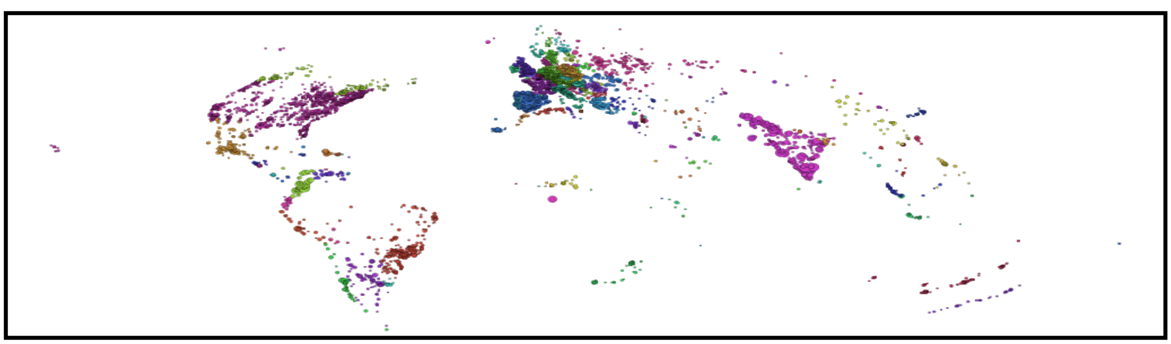

Figure 1. Geographic location of participating students

Additionally, students participating from different countries performed at different levels and spent varying amounts of time on the site. Table 1 illustrates the variation in participation (overall registration as well time spent on one of the website components—the homework problems) and in performance (number of certificate earners and average points earned out of 100) for the top 12 countries represented. Points, here, are out of 100, indicating the grade for the whole course. The mean performance includes all students in that country; points distributions were highly skewed upwards.

Table 1. Top 12 countries by participation, performance metrics

Country	Number of registrants	Number of certificate earners	Mean/SD total points for all registrants	Mean/SD hours spent on homework
United States	26309	1321	5.65 (19.33)	1.57 (5.24)
India	13044	838	7.84 (22.11)	1.56 (4.49)
United Kingdom	8430	550	7.25 (21.66)	2.07 (6.02)
Colombia	5955	458	8.93 (23.36)	2.56 (6.57)
Pakistan	4308	212	6.50 (19.36)	1.32 (4.24)
Brazil	3852	190	5.5 (18.99)	1.40 (4.72)
Spain	3684	535	14.43 (30.29)	3.25 (6.95)
Mexico	2883	150	6.28 (19.71)	1.72 (4.86)
Canada	2834	162	6.46 (20.65)	1.71 (5.45)
Russian Federation	2029	195	10.66 (26.70)	2.39 (5.73)
Poland	1392	187	14.49 (29.50)	3.18 (6.26)
Greece	1386	187	13.67 (29.73)	3.68 (8.65)

3.2. Variability in offline collaboration

In 6.002x, there was a range of student responses to the question of working with collaborators offline. While most respondents reported that they worked on 6.002x completely on their own (75.71%), a notable portion of students reported that they worked offline with another 6.002x student (17.68%) or that they worked with someone who has expertise in the content area (2.57%).

Table 2. Proportion of respondents working with collaborators offline

Working offline	Count	Percent of respondents to this question
I worked completely on my own	2359	75.71
I worked with another person who is also completing the course.	551	17.68
I worked with someone who teaches or has expertise in this area.	80	2.57
Other	126	4.04

We then conducted a one-way analysis of variance (ANOVA) of respondents' final grade in the course by the type of offline collaboration they reported. The ANOVA test showed that there were significant differences in grade by students' collaboration with others offline: $F(3, 3075) = 14.28, p < 0.01$. (We report both our F-statistic and p-values for each ANOVA. With the F-test, we also note the degrees of freedom based on the number of groups for each categorical survey question as well as the overall sample size for the subset of students who received that question.) Figure 2 shows variation in performance by respondents' collaboration with others offline.

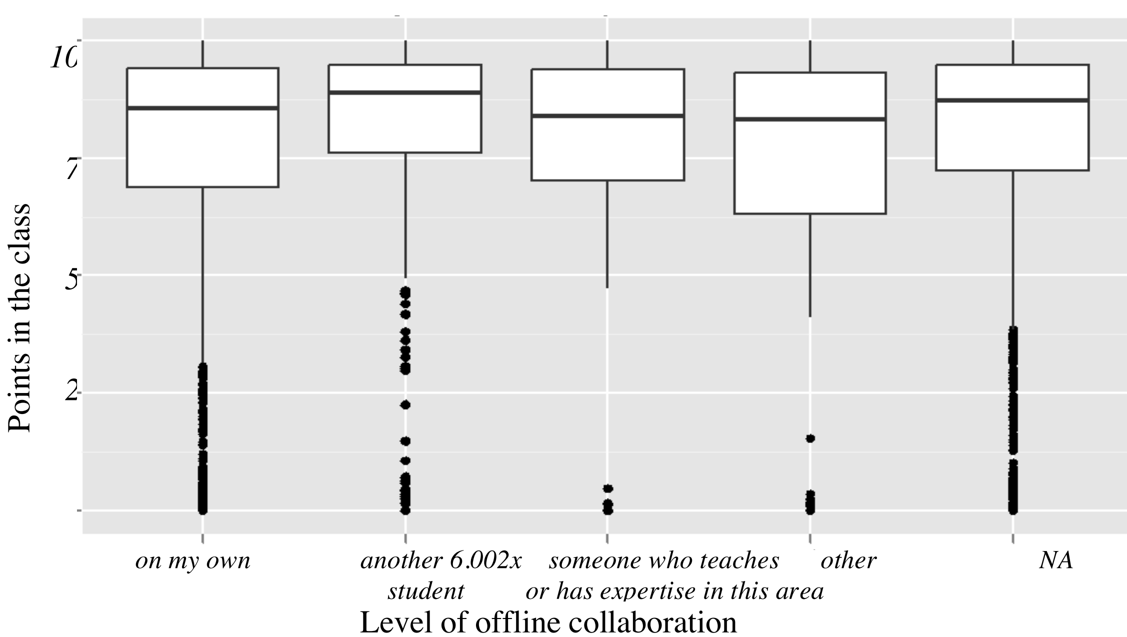

Figure 2. Performance by collaboration offline

3.3. Educational attainment

The end-of-course survey posed the question, "What is the highest degree you have completed?" Of the 2,314 responses, the highest percentage of participants (36.63%) reported having a bachelor's degree, followed by a master's or professional degree (27.87%), and a secondary or high school degree (26.68%). Although a statement on the course site recommended that students have prerequisite knowledge in advanced academic coursework such as an Advanced Placement (AP) level physics course in electricity and magnetism, there were still 72 of the surveyed students who reported having only attained elementary/primary school or junior secondary/high school level education, an environment in which advanced level courses were not likely to be offered. The table below shows the highest degree earned by all students who completed the end-of-course survey.

Table 3. Proportion of respondents by educational attainment levels

Degree level	Count	Percent of respondents to this question
Elementary/primary school	16	0.53
Junior secondary/high school	56	1.86
Secondary/high school	804	26.68
Bachelor's degree	1104	36.63
Master's or professional degree	840	27.87
PhD in a science or engineering field	178	5.91
PhD in another field	16	0.53

We then compared the overall course achievement, measured by points earned in the course, of students with the various levels of educational preparation. A one-way analysis of variance showed that there were significant differences between the mean scores of groups with different levels of preparation, $F\ (6,\ 2966) = 10.20,\ p < 0.01$. The group of students with the highest mean for points earned in the class was comprised of those who reported having a PhD in a science or engineering field. As might be predicted, the group with the lowest mean points was comprised of students who reported having only a primary/secondary degree prior to enrollment. It should be noted, however, that although the mean score of this group was lowest, the range of scores shows that there were individual students who performed very well. There was little difference in the mean scores of groups who reported having a bachelor's degree, junior secondary/high school, or secondary/high school levels of preparation. The figure below illustrates the total course points earned by students with varying levels of educational preparation.

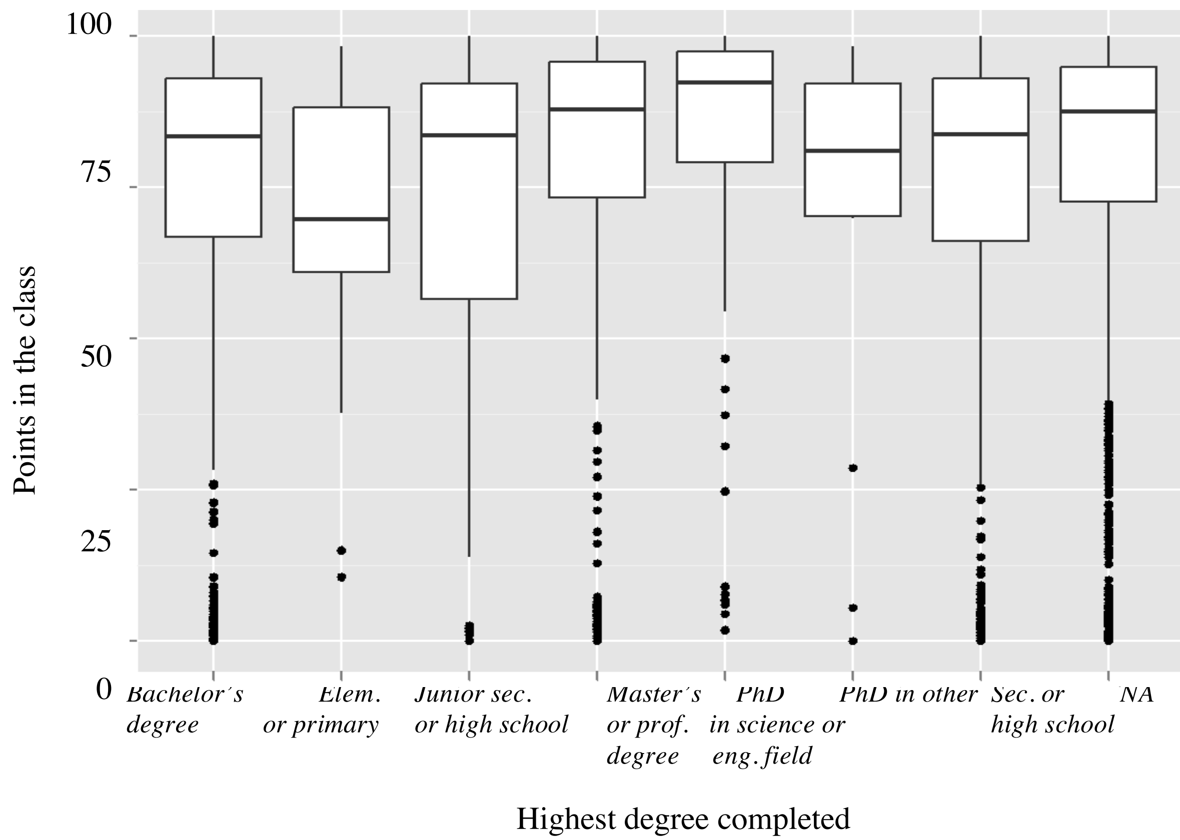

Figure 3. Performance by level of educational attainment

3.4. Content familiarity

In order to better understand the 6.002x learners' knowledge about electrical engineering prior to taking the course, we placed a question on the post-course survey that asked if they taught electrical engineering in any setting. As evidenced by the table below, the majority of survey respondents (86.61%) were not involved in teaching this subject. Two hundred and fifty-five respondents (8.80%) reported teaching electrical engineering in a college or university setting, whereas fewer respondents reported teaching the subject in other settings (2.59%) or at the high school or secondary level (2.00%). This small but important population is noted in other open online courses [3].

Table 4. Proportion of respondents by experience teaching content

Teaching status	Count	Percent of respondents to this question
I do not teach EE.	2510	86.61
I teach EE elsewhere.	75	2.59
I teach EE in college/university.	255	8.80
I teach EE in high school/secondary level.	58	2.00

An analysis of variance showed that there were no significant mean differences in overall course achievement when groups were compared using reported teaching status as the grouping factor $F(3, 2857) = 1.54, p = 0.20$. The mean course points for 6.002x students who reported teaching electrical engineering in high school or secondary school was lower than the mean for the other groups. The group who reported teaching electrical engineering in a college or university had the highest mean course points, although their scores were not significantly different from the mean of those who did not teach or reported teaching elsewhere. The figure below illustrates the total course points earned by students reporting teaching or not teaching electrical engineering.

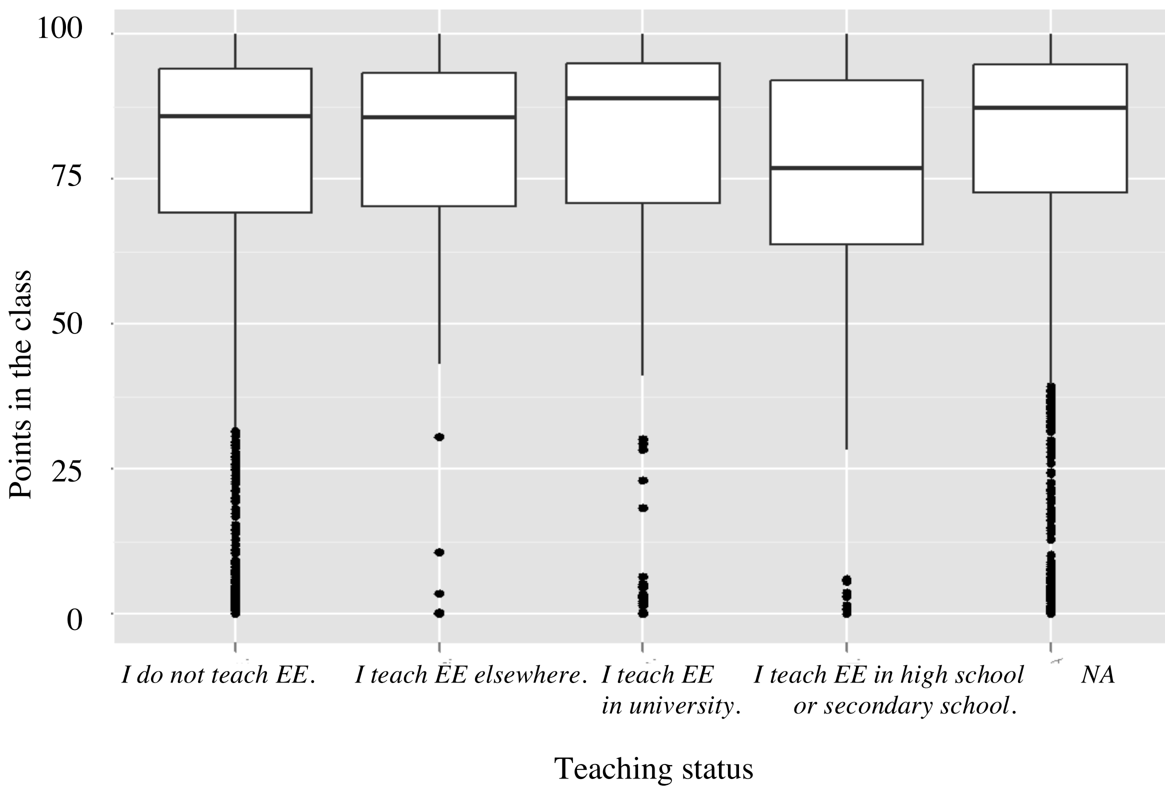

Figure 4. Performance by teaching content

3.5. Motivation for enrollment

Of substantial interest to MOOC developers and researchers trying to understand completion rates are reasons why students enroll in the course. This initial motivation for enrollment may aid in predicting how much effort students will exert in the course. For 6.002x, the most frequent response (55.41%) to the question, "What is your primary motivation for taking 6.00x?" indicated that student enrollment was driven by a desire to gain knowledge and skills, followed by a desire for personal challenge (25.58%). The option that received the fewest responses was related to gaining a social understanding and friends (0.43%). It should be noted, however, that students were limited to one selection for this question, thus limiting the responses to their primary reason. Students may have had a combination of reasons for enrolling. The table below shows the proportion of all response to this survey question.

Table 5. Proportion of respondents by reason enrolled in class

Reason given	Count	Percent of respondents to this question
Employment/job advancement opportunities	97	8.27
Other	40	3.41
Preparation for advanced standing exam	28	2.39
Social understanding and friends gained as a result of taking the course	5	0.43
The entertainment value of the course	53	4.52
The knowledge and skills gained as a result from taking the course	650	55.41
The personal challenge	300	25.58

An analysis of variance showed that there were marginal significant mean differences in overall course achievement when groups were compared using reason for enrollment as the grouping factor $F(6, 1155) = 2.07, p = 0.05$. The mean course points from students responding "knowledge and skills gained" and "personal challenge," reasons that may be interpreted as conveying intrinsic motivation to learn, were not significantly different from the means of groups responding "employment/job advancement opportunities" or "preparation for advanced standing exam", which may be interpreted as extrinsic sources of motivation. The figure below illustrates the total course points earned by groups by reason for course enrollment.

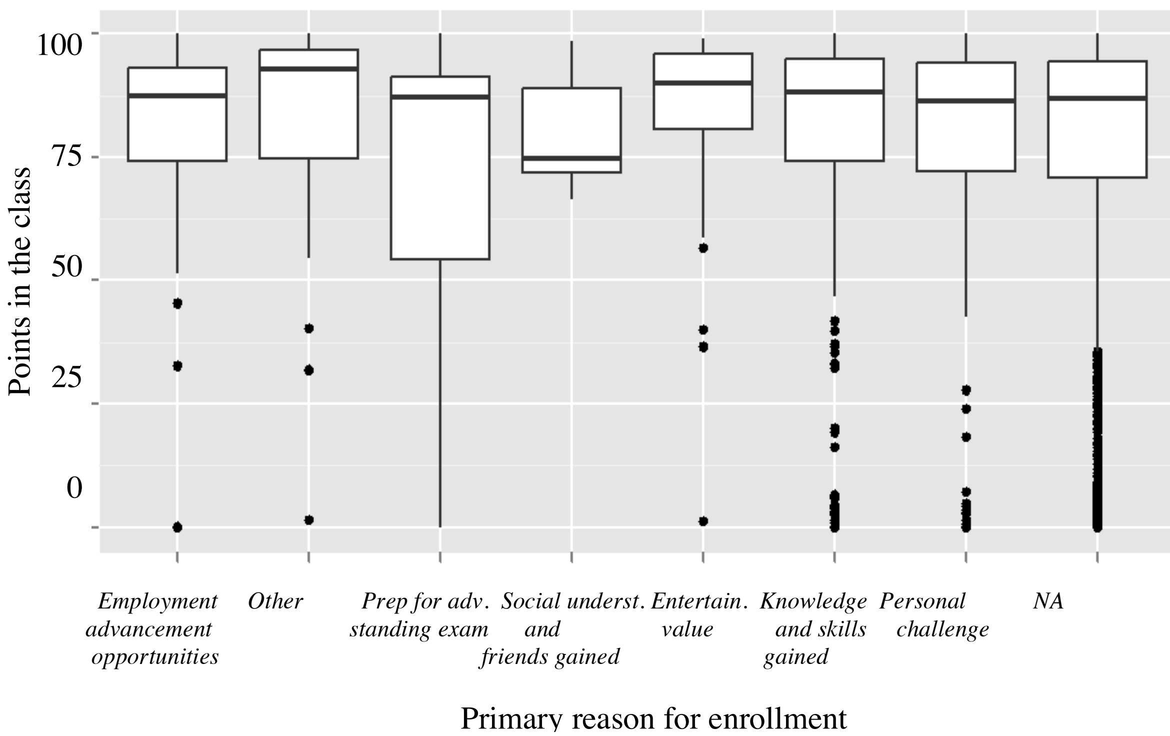

Figure 5. Performance by reason enrolled in class

4. Implications

We demonstrate that there is a high degree of variability in the backgrounds and behaviors of students in the first MIT MOOC class. Some of these characteristics also relate to notable differences in performance as well. Our descriptive work in this study serves as a spotlight on the important challenges that such a diverse classroom poses for instructors, course designers, and education researchers. In further studies, we delve more deeply into the predictors of student achievement.

As illustrated by points earned in 6.002x by students with various levels of preparation, prior educational experience was an important factor in predicting student success. Students who are less prepared may need experiences that scaffold their understanding of pivotal course concepts. It must be acknowledged that points in the class may also be a reflection of time and effort expended by students and not necessarily an indication of an increase in knowledge or skills in this content domain. However, the lower performance of 6.002x students who come in with lower educational attainment suggests that further exploration into the needs of this particular group may be warranted. Another group of participants whom we identified in this analysis was comprised of those who teach at the high school or secondary school level. MOOCs are an ideal mechanism for delivering continuing education to interested individuals, and secondary school educators may be a prime audience for this type of learning experience. It is important to note, though, that this is the first edX class, and the generalizability of results in this dynamic, early stage of MOOCs has yet to be determined.

In this study, we note that students who collaborate with others offline may do better in the class. This suggests to MOOC providers that supporting different venues for student-student interaction may help learning in MOOCs. The marginal significance of students' reasons for enrolling is also provocative. Even among the limited sample of students who continued to participate in the class through the end and who responded to the exit survey, there is a marginally significant difference in performance between students who enrolled for different reasons. MOOC providers may be able to differentially support students who register for classes for different reasons (e.g., [4]). Knowledge of prior student experiences may be a helpful piece of information for MOOC providers to support unique individuals in these new massive classrooms. In future work, we investigate the complex multivariate relationships between prerequisite knowledge, different educational experiences, and other "background" variables that characterize different groups of students, and success in MOOCS. MOOCs must clearly understand the types of students they are targeting as well as the prerequisite knowledge and experiences necessary to succeed in MOOC classes.

5. Acknowledgements

We thank MITx for data access, and MIT Professor David Pritchard and members of his RELATE (Research in Learning, Assessing, and Tutoring) group for their work in data preparation and subsequent analysis.

6. References

[1] Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Review of Educational Research, 69(1), 21–51.

[2] Stump, G.S., Hilpert, J., Husman, J., Chung, W.T., & Kim, W. (2011). Collaborative learning in engineering students: Gender and achievement. Journal of Engineering Education, 100(3), 475-497.

[3] Fredericks, C., Rayyan, S., Teodorescu, R., Balint, T., Seaton, D., & Pritchard, D.E. (2013). From Flipped to Open Instruction: The Mechanics Online Course. LINC 2013, Boston, MA.

[4] Kizilcec, R.F., Piech, C., & Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. LAK'13 Leuven, Belgium.