Pre-registered Analysis Plan for

“Early Interventions in Online Education”

René F. Kizilcec[1]

Justin Reich[2]

Yigal Rosen, Dustin Tingley, Selen Turkay, Joseph Williams & Michael Yeomans[3]

1 Stanford University

2 Massachusetts Institute of Technology

3 Harvard University

Table of Contents

Introduction and Motivation

Over 35 million people have enrolled in massive open online courses (MOOCs) between 2012-2015 (Shah, 2015). People enroll in MOOCs for a variety of reasons and many never intend to complete the course (Kizilcec, Piech & Schneider, 2013; Kizilcec & Schneider, 2015; Reich, 2014). As a result, though enrollment in MOOCs has exploded, course completion rates still lag well behind those in traditional education (Banerjee & Duflo, 2014; Ho et al., 2015). Yet even learners who explicitly intend to finish a class struggle to reach their goal. This low retention rate curtails the full benefits that MOOCs can provide, but it is a problem that may be addressed. Research in the behavioral sciences and psychology suggests that small changes in the design of the learning environment can help people follow through on their goals (Thaler & Sunstein, 2008; Walton, 2014).

In this research, we focus on how simple interventions at the start of an online course can foster academic achievement. Many learners who disengage from a MOOC do so early in the course, thus early interventions can support them while their commitment is still strong. The interventions are short, self-contained writing activities, so they are easily implemented by course designers or platform administrators. Previous research has shown positive effects of these interventions in studies with a small number of courses (Kizilcec, Saltarelli, Reich, & Cohen, 2017; Kizilcec, Davis, & Cohen, 2017; Yeomans & Reich, 2017). Our goal in this research is to systematically evaluate these interventions, individually and in combination, in a large-scale experiment across many courses and many institutions. To our knowledge, this will be among the largest intervention experiments in the history of online education. The scope of this research provides a unique opportunity to answer important questions on how to support a diverse population in their pursuit of long-term educational goals.

Experimental Design

Most MOOCs have a course survey embedded at the beginning of the class material (e.g., in edX, as "chapter zero" of the course), which includes questions about demographics and other covariates. We will standardize the pre-course survey questions across institutions and courses to compare treatment effects across diverse samples, and to identify whether some interventions are more effective amongst well-defined subgroups of learners. We will add the interventions at the end of this survey and assign each learners to one of several conditions. Based on our previous research, we expect that a substantial fraction of learners will not engage with the intervention - either skipping the writing prompt entirely, or else quitting the course survey to begin the course. However, we will conduct all our analyses at the level of intent to treat, estimated among all learners who were *exposed* to the intervention, regardless of their level of engagement.

Assignment to conditions will be stratified random to balance four pre-course variables that have been strong predictors of course completion in previous research (Kizilcec & Halawa, 2015). The four variables used to stratify assignment will be: *intentions to complete the course assessments [all or most; few or none (or blank)]*; *intended hours spent on the course per week [0-5 (or blank), 6+]; previous MOOCs completed [0 (or blank),1-3,4+];* and *previous education [graduate degree, bachelor's degree, less than bachelors (or blank)]*. This stratified assignment increases the power of our design, both for average treatment effect estimates and for subgroup analysis. Note that conditional on this stratification, assignment to conditions will be completely random. However, it is possible that our random assignment will be unbalanced across covariates that are not explicitly part of our stratification. We will test the randomization balance ex post across measurable covariates, and if this randomization is demonstrably unsuccessful, we will also conduct analyses that control for those unbalanced covariates to preserve causal inference across conditions.

Intervention Design

The interventions in this experiment have several important features in common. They are self-contained writing activities that require no follow-up from the experimenters. They take only minutes to complete and uninterested learners can easily opt out. Most importantly, the interventions have been successfully tested in previous randomized control trials in MOOCs. We evaluate two types of intervention - *affirmation* and *plan-making* - in a 2x3 design, for a total of five intervention conditions and one pure control. The affirmation factor will have two levels - control (i.e. no affirmation) and affirmation. The plan-making factor has three levels - control (i.e. no plan-making), short plans, and long plans. In the two conditions that include multiple interventions, the order of the interventions will always put affirmation first, and plan-making second.

The affirmation intervention aims to reduce a social psychological barrier that can hinder learners from achieving their potential in a learning environment. Learners can be concerned about their belonging in the MOOC and worry that others could see them as less capable because of one of their social identities (e.g., nationality, gender, social class). In MOOCs offered by elite universities, learners from underprivileged backgrounds may doubt their “fit” into the archetypal role of a high-achieving student at such an institution. The value relevance affirmation draws on two established intervention paradigms: values affirmation based on self-affirmation theory (Cohen, Garcia, Apfel, & Master, 2006; Steele, 1988) and a relevance intervention to foster connections between course materials and learners' life (Hulleman & Harackiewicz, 2009). Learners select 2-3 values or qualities from a list that are most important to them. They then write about how takings this online course reflects and serves their cherished values. To support internalization, learners then write a note to their future self to remind them of the importance of their values in the course. Previous research (Kizilcec, Saltarelli, Reich, & Cohen, 2017) tested this intervention in two MOOCs which drew from many different countries, and showed that the course completion rates among people from developing countries roughly doubled, while only slightly reducing completion rates in developed countries. Moreover, treated learners were more likely to sign up for future courses, indicating lasting effects to close the global achievement gap. Another experiment in a Chinese MOOC found that a value relevance affirmation, adapted for the interdependent cultural context in China, increased completion rates by 41% among the most identity-threatened subgroup of learners this a language course, lower-class men (Kizilcec, Davis, & Cohen, 2017).

The second intervention factor address an important strategy to succeed in online learning: strategic planning (Kizilcec, Perez-Sanagustin, & Maldonado, 2016). In particular, people tend to ignore the procedural details of goal pursuit, naively thinking that their strong intentions will sustain over time. The planning prompt intervention asks learners to carefully consider those details while their intentions are strong, including specific times, locations, and study strategies. Learners are also encouraged to write down their self-generated plans as a reminder. Previous research (Yeomans & Reich, 2017) has shown that planning prompts in three MOOCs increased course completion by 29%, and also increases learners' willingness to upgrade to verified certificates during the course by 40%. In this experiment, we will include two different plan-making interventions: *long plans*, which ask learners to plan their participation over the entire course; and *short plans*, which ask learners to plan their participation over the first two weeks of the course. The short plans intervention might allow learners to set sub-goals and write more concrete plans, which might then induce habit formation or other long-term behaviors that are consistent with their initial intentions.

Outcome Measures

The EdX and Open EdX platforms passively log all course activity, which allows us to directly measure many outcomes for learners in the course. Our primary focus is to test whether the interventions help learners complete the course. We therefore assess whether a learner has achieved a high enough grade in the course to earn a "basic" certificate or statement of accomplishment.

We note that EdX (the platform for MIT and Harvard courses, but not Stanford courses) recently changed its certification policy, such that learners cannot achieve an official designation of course completion without paying for the "verified track" in the course. Before 2016, learners could earn a certificate for free, but these basic certificates no longer exist for courses at MIT and Harvard. In contrast, Stanford courses in this research do not offer paid "verified" certificates, only free basic certificates. For consistency across courses on different platforms, we will determine from the course tracking logs for Harvard and MIT courses whether a learner achieved a high enough grade to earn a "verified" certificate, even if the learner did not pay[1].

- Learner earned a basic certificate (or equivalent)

Due to this variation course structure, we also want a secondary proxy for course progress, that could be measured across all certificate types, at all schools. To do this, we decided to also count the percentage of each course's videos that each learner begins to watch, according to the tracking logs.

- Percentage of course lectures that a learner began to watch

Second, we will assess two additional outcomes related to whether learners pay to enter the "verified" track of a course for an opportunity to earn a "verified certificate". This option is only available in Harvard and MIT courses and learners can pay to upgrade to the verified track before or during the course (including after taking the survey).

- Learner earned a verified certificate (regardless of purchase time)
- Learner upgraded to the verified track after the survey (regardless of completion)

Third, we will assess an intermediate self-report outcome that measures the immediate effects of the interventions on learners' expectations of success. At the end of the course survey,

[1] There is some uncertainty about how this measure is calculated. Our approach is to use grades for learners who are not on the verified track and apply the same performance threshold as on the verified track (e.g., grade above 70%) to determine if a basic certificate was achieved in theory.

after the intervention activities, we included a question asking learners to predict the percentage likelihood that they complete the course.

- Self-reported percentage likelihood of completing the course

Finally, an outcome that we are interested in evaluating at a later point when it becomes available is whether the interventions affected learners' future enrollment in other courses at the same institutions. Enrollment records will be accessible by the course team for all Harvard, MIT and Stanford courses (but not EdX courses from other schools). We aim to assess this outcome on January 1, 2018.

Sample Determination / Exclusions

The intervention was included in every class offered by Stanford, Harvard, and MIT on the (Open) EdX platform with a start date after Sept 1, 2016 and an end date before March 1, 2017. We lists all of these courses in a separate file, *courselist.csv*, along with class-level details (length, enrollment type, grading scheme, etc.). Some courses at these institutions that started in this timeframe were excluded due to administrative hurdles or because the course format differed substantially from the other courses (e.g., a course that lasts just one week). Although we list the specific courses included in our dataset, we cannot determine our sample size in advance because learners enroll on their own volition.

The survey software (Qualtrics) records every learner who opens the course survey. Although many learners who enroll in the course do not take the survey, prior work shows that the response rate is much higher among more active and committed learners (Reich, 2014). Our analyses focus on a subset of learners who start the survey, defined by criteria that focus our investigation on learners expected to benefit from the interventions, based on prior intervention research. First, some learners do not progress far enough in the survey to reach the point at which they are randomized into a treatment group and exposed to an intervention activity; these learners will be excluded. Second, some learners take multiple courses covered by our experiment, but we only analyze learners' first exposure and exclude those who sign up for multiple courses in a thirty-day window to avoid exposure to different treatments. Third, most courses allow learners to enrol at any time, and complete the course survey whenever they want. However, most dropouts occur early, and those who take the survey mid-course may not be affected by the intervention; and learners who join late may not be able to finish the course due to structural reasons such as deadlines and closing dates.

In summary, all our analyses will only focus on learners who:

(a) progressed far enough in the survey to be assigned to conditions and exposed,

(b) were not exposed again within 29 days following their initial exposure,

(c) started the survey within the first hour of accessing any course materials,

(d1) started the course within 14 days of the start of a cohort course, OR

(d2) start the course more than 30 days before the end of a self-paced course.

In prior plan-making studies, our pre-registered analysis plan stated that we would target learners who were fluent in written English, so that their plans could be reported naturally. Most non-fluent learners skipped the text questions in the survey, or wrote in another language, and in fact they had a somewhat smaller treatment effect than our focal sample. We had also stated a focus on learners who intended to complete the course (cf. Sheeran & Gollwitzer, 2005). In that sample we found no moderation of plan-making by intentions. However, we still believe that intentions could be an important moderator in the new sample. We therefore conduct the primary analyses for the planning intervention using the subpopulation of fluent English speakers who intend to complete all assessments.

Hypotheses

The main analysis in this project concerns the estimation of treatment effects. For the plan-making intervention, we are interested in the individual treatment effects of the two different versions of plan-making (short and long), as well as the average effect across both versions of plan-making, compared to the no-plans control conditions.

For the affirmation intervention, we are interested in treatment effects on specific groups of at-risk learners, primarily learners in less developed countries (HDI < 0.7), whereas the intervention is not expected to benefit learners who are not at-risk, such as those in more developed countries (HDI > 0.7). We fit a model with first- and second-order effects of affirmation condition and learner subgroup (e.g., with *affirm*: 1=yes, 0=no; *highHDI*: 1=yes, 0=no; and a *affirm * highHDI* interaction). This allows us to identify the treatment effect for at-risk learners (coefficient on *affirm*), the achievement gap between at-risk and not at-risk learners (coefficient on e.g. *highHDI*), and how the treatment effect differed for not at-risk learners (coefficient on interaction, e.g. *affirm * highHDI*). The primary analysis focuses on at-risk learners in terms of HDI as defined above, replicating prior work (Kizilcec et al., 2017). Secondary analyses explore moderators of this effect in terms of national social identity threat (the intervention should particularly help those who experience identity threat based on their nationality) and English language proficiency (those who are not proficient may experience low belonging and should benefit from the intervention more). Additional secondary analyses examine three other theory-based definitions of at-risk learners: learners with low socioeconomic status, women in male-dominated (<20% women; cf. Walton et al., 2015) course, and underrepresented minorities in the United States.

Analysis

This pre-registration includes three R scripts that lay out details of the dataset and planned confirmatory and planned exploratory analysis. First, the `sampledata.R` file generates a "toy" dataset (`simdat.rda`) that resembles the dataset that will be used in the analysis. The details of creating this dataset from each course are idiosyncratic to each university (and some courses may require different protocols), and we will strive to report those ex post where possible.

Second, the `prereg.R` file lays out how the subpopulation of interest is determined (i.e. exclusions) and what analyses will be performed. The analyses are organized based on their priority, starting with models that essentially replicate analyses from prior work (primary analyses), and continuing with additional theory-driven models (secondary analyses). The analyses are also organized based on the priority of the outcome measures and subsamples, as defined above. The analysis strategy is to evaluate treatment effects using hierarchical regression models. Institution, course, and strata will be modeled as random intercepts and the four individual stratification variables will serve as covariates (see details in script).

Third, the `sparsereg.R` file lays out the planned procedure for investigating heterogeneous treatment effects using a data-driven approach. We fit Bayesian LASSOplus models (implemented in the *sparsereg* R package) to discover interactions between the treatment and learner characteristics collected prior to exposure to intervention activities, with the goal of discovering an optimal treatment assignment for future learners.

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web,* 687-698.

Banerjee, A. V., & Duflo, E. (2014). (Dis) organization and Success in an Economics MOOC. The American Economic Review, 104(5), 514-518.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *science*, *313*(5791), 1307-1310.

Ho, A. D., Chuang, I., Reich, J., Coleman, C. A., Whitehill, J., Northcutt, C. G., ... & Petersen, R. (2015). *HarvardX and MITx: Two Years of Open Online Courses Fall 2012-Summer 2014.*

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. Science, 326(5958), 1410-1412.

Kizilcec, R. F., Davis, G. M., & Cohen, G. L. (2017). Towards equal opportunities in MOOCs: Affirmation reduces gender & social-class achievement gaps in China. In *Proceedings of the Fourth ACM Conference on Learning @ Scale*.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education, 104*, 18-33.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge,* 170-179.

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. Transactions on Human-Computer Interaction 22, 2.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *In Proceedings of the Second ACM Conference on Learning@ Scale,* 57-66.

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

Reich, J. (2014). MOOC completion and retention in the context of student intent. *EDUCAUSE Review Online.*

Dhawal S. (2015). MOOCs in 2015: Breaking Down the Numbers. *EdSurge. Accessed online at https://www.edsurge.com/news/2015-12-28-moocs-in-2015-breaking-down-the-numbers*

Sunstein, C., & Thaler, R. (2008). *Nudge. The politics of libertarian paternalism*. New Haven.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73-82.

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J. & Zanna, M. P. (2015). Two brief interventions to mitigate a 'chilly climate' transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology,* 107**,** 468–485.

Yeomans, M. & Reich, J. (2017). Planning Prompts increase follow-through in MOOCs. W*orking paper.*