

A Randomized Experiment Testing the Efficacy of a Scheduling Nudge in a Massive Open Online Course (MOOC)

Rachel Baker

University of California, Irvine

Brent Evans

Vanderbilt University

Thomas Dee

Stanford University

An increasing number of students are taking classes offered online through open-access platforms; however, the vast majority of students who start these classes do not finish. The incongruence of student intentions and subsequent engagement suggests that self-control is a major contributor to this stark lack of persistence. This study presents the results of a large-scale field experiment (N = 18,043) that examines the effects of a self-directed scheduling nudge designed to promote student persistence in a massive open online course. We find that random assignment to treatment had no effects on near-term engagement and weakly significant negative effects on longer-term course engagement, persistence, and performance. Interestingly, these negative effects are highly concentrated in two groups of students: those who registered close to the first day of class and those with .edu e-mail addresses. We consider several explanations for these findings and conclude that theoretically motivated interventions may interact with the diverse motivations of individual students in possibly unintended ways.

Keywords: *educational policy, experimental research, higher education, postsecondary education, technology*

MOTIVATED by concerns about low persistence in online higher education, researchers and policy makers have called for strategies to increase student engagement in online courses. This paper examines the effects of an inexpensive, scalable, easily implemented, and theoretically grounded intervention aimed to increase persistence. On the first day of a science-based massive open online course (MOOC), a randomly selected subset of students was offered the opportunity to schedule when to watch the first lecture video. This scheduling nudge, grounded in recent work on commitment devices from behavioral economics, provided students with the opportunity to create structure within the online course as a way to guide their future behavior.

Three stylized facts motivate this study. (a) Postsecondary students are increasingly earning college credit from multiple sources and institutions: traditional 2- and 4-year schools and in-person and online courses offered both through accredited universities and through open-access platforms (including MOOCs; Andrews, Li, & Lovenheim, 2014; McCormick, 2003). (b) Whereas rates of college attendance have increased dramatically over the past 40 years, rates of persistence and completion have not shown a similar rise, especially disadvantaging those from lower socioeconomic backgrounds (Bailey & Dynarski, 2011).

(c) As more classes become asynchronous and increasing numbers of students attend school part-time, time management and other meta-academic skills have become an important contributor to low levels of completion (Hill, 2002; Roper, 2007).

Online classes have become an important part of the higher education landscape and open-access courses provided by third-party nontraditional sources are a large and growing component of this phenomenon. Over the last decade, MOOCs have grown from a few courses offered by individual instructors to an enormous enterprise incorporating hundreds of the world's top universities and multiple platforms. With tens of thousands of students enrolling in each of the hundreds of courses available, tens of millions of students have accessed MOOCs globally. Although the majority of MOOC users are not seeking official college credit, there are several pathways through which students can receive college credit or other valuable certifications. Udacity offers nanodegrees and Coursera offers specializations, both of which provide a program of study combining several courses with the potential to earn certification. Arizona State University (ASU) provides general education credits for completing ASU sponsored MOOCs (Haynie, 2015).



As MOOCs have risen in prominence, scale, and scope, and as their role in postsecondary credentialing increases, there has been widely publicized descriptive evidence that surprisingly large numbers of registrants fail to finish these courses. Only slightly more than half of MOOC registrants typically watch the first lecture of their course, and lecture watching declines continuously as a course progresses such that only 10% of registrants watch 80% or more of a course's lectures (Evans, Baker, & Dee, 2016; Jordan, 2014; Perna et al., 2015).

Although a number of explanations can reasonably be applied to these low levels of persistence (e.g., students purposefully sample a range of different courses, students wish to focus on a select topic within the course), in this study we focus on one factor: the advanced time management skills needed to persist in these unstructured environments. As online classes are often offered asynchronously, success in these classes requires a set of skills and tools other than persistence in an in-person class. Class and study time are not set, so if students do not schedule time in advance, more immediate tasks can take priority over completing course work. Thus, one candidate explanation for the sharp contrast between the volume of course registrants and the pattern of continuously declining student persistence is that a sizable number of MOOC students have time-inconsistent preferences. These registrants may genuinely desire to engage with their selected course when they register but subsequently choose not to engage when the opportunity actually arises (i.e., when weekly lectures are made available). Consistent with this hypothesis, we find that the modal student watches lecture videos around 8:00 p.m. local time, which suggests students face trade-offs when watching lectures during leisure hours. Motivated by this conjecture about the possible role of self-control and the lack of established study time, this study presents the results of a field experiment in which over 18,000 students in an introductory science-based MOOC were randomly assigned to an opportunity to commit to watching a lecture video at a specific time and date of their own choosing.¹

Due to their large, diverse enrollments, MOOCs provide a fertile testing ground for interventions that may prove useful in more traditional learning settings. At the same time, the particular challenges present in the online context (such as less structure and asynchronicity) call for the examination of new, specific interventions for online classes. Interventions attempting to improve engagement and persistence in traditional higher education are common (e.g., tutoring, study skills classes, and orientation programs), but the design and delivery of such supports may need to be modified for MOOCs. Although some may argue such supports reduce the signaling value of the credential, they likely enhance human capital if they increase educational engagement, persistence, and success.

The results from our field experiment are surprising. Suggesting that students schedule time to watch the first

lecture video in each of the first 2 weeks of the course does not affect near-term measures of engagement. The opportunity to schedule does have weakly significant negative effects on several longer-run course outcomes, including the total number of lectures watched in the course, performance as measured by final grade, and whether the student earned a completion certificate. These negative effects are strongest among students with .edu e-mail addresses and those who registered for the course shortly before it began.

We first describe the relevant bodies of literature that contribute to the design of this study before describing the experimental setting, procedures, and results in more detail.

Prior Literature

Our randomized experiment is motivated by three disparate strands of prior literature: the growing literature on persistence in MOOCs, the higher education literature on time management and scheduling study time, and the behavioral economics literature on precommitment devices. We discuss each in turn below.

Persistence and Performance in MOOCs

Understanding persistence in online higher education has become an increasingly important undertaking given the rapid growth of both credit- and non-credit-bearing online courses, the documented low persistence rates in these courses, and the institutional goals of MOOCs to simultaneously improve educational outcomes and lower educational costs (Evans & Baker, 2016; Hollands & Tirthali, 2014). It is thus important to unpack the correlates of enrollment and persistence for students in these courses.

Evans et al. (2016) document course-, lecture-, and student-level characteristics that are related to engagement and persistence within individual MOOCs. Additionally, several efforts have examined how fine-grained measures of student behavior in a MOOC are related to persistence and performance (e.g., Halawa, Green, & Mitchell, 2014), including temporal elements, such as when students watched a lecture or completed a quiz relative to its release (e.g., Ye et al., 2015).

However, no prior studies have examined this question causally. That is, no studies to date have examined whether the specific behaviors within a course, such as if students schedule when to watch lectures, can be affected, or if such interventions have an effect on course outcomes. Of the extant literature, there are two MOOC studies that are most closely aligned with this current paper. Banerjee and Duflo (2014) compared the persistence and performance of students who registered just before and just after the stated (but not imposed) enrollment deadline. They found students who registered later have worse course outcomes and offer the explanation that these students, who exhibited a tendency

toward procrastination, may have had a hard time meeting course deadlines. The study does not attempt to influence student behavior. Zhang, Allon, and Van Mieghem (2015) encouraged a random subset of students in one MOOC to use the discussion forum in an effort to increase social interaction. They found that although the nudge increased the probability of completing a quiz, it had less of an effect on student performance. Our study similarly employed a randomized control trial using a nudge intervention to adjust students' behavior in the MOOC. Instead of encouraging discussion board interaction, we attempted to alter time management skills by suggesting students schedule a time to engage with the course material. This enables the measurement of the effect of this scheduling suggestion on course engagement and persistence in the absence of other confounders.

Time Management and Scheduling Study Time

Prior research has repeatedly demonstrated that time management is an important skill related to college performance in traditional higher education settings. Planning ahead to study course materials throughout the term, as opposed to cramming right before a deadline, is positively correlated with a higher college grade point average (GPA; Hartwig & Dunlosky, 2012). Similarly, Macan, Shahani, Dipboye, and Phillips (1990) found that scores on a robust time management scale were positively related not only to higher college GPA but also to students' self-perceptions of performance and general satisfaction with life. College students with better time management skills not only scored higher on cognitive tests but were actually more efficient students, spending less total time studying (Van Den Hurk, 2006).

There is not a large literature focusing explicitly on the scheduling component of time management. However, short-range planning, including scheduling study time, was found to be more predictive of college grades than SAT scores (Britton & Tesser, 1991). Misra and McKean (2000) offer one potential mechanism for these results: Time management is an effective strategy to reduce academic stress and anxiety, which in turn may increase performance.

Important for our specific context, these results have been shown to extend to online learning settings. Student success in online courses in higher education is negatively related to procrastination (Elvers, Polzella, & Graetz, 2003; Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011). In a study of online learners who completed degrees, students identified that developing a time management strategy was critical to their success (Roper, 2007). This has also been found to be true in the context of MOOCs: Using a survey of MOOC dropouts, Nawrot and Doucet (2014) found a lack of quality time management to be the main reason for withdrawing from the MOOC. Indeed, a sizable portion of MOOC learners have flexible professional schedules and state that they

take MOOCs due to course flexibility (Glass, Shiokawa-Baklan, & Saltarelli, 2016). In fact, Guàrdia, Maina, and Sangra (2013) argue providing a scheduling structure with clear tasks is one of 10 critical design principles for designing successful MOOCs.

The goal of our study, unlike many of the previously discussed works, is not to survey students about their study strategies to look for a relationship between study skills and academic outcomes. Rather, given the consistent evidence that good time management practices are associated with positive outcomes, we attempted to nudge students to schedule study times. Song, Singleton, Hill, and Koh (2004) and Nawrot and Doucet (2014) directly called for such interventions targeting the development of time management strategies, and the scheduling device tested in our study is one such intervention. A work-in-progress paper presented at the Learning @ Scale conference provides a similar test by randomly informing a small ($N=653$) sample of MOOC users that a set of study skills has been reported as effective (Kizilcec, Pérez-Sanagustín, & Maldonado, 2016). That study showed no effect on engagement and persistence outcomes, but like all other previous studies, it did not have the power to identify small effects. Our experiment focuses on a scheduling prompt as opposed to providing information on study skills and is much more highly powered.

Precommitment

The economic literature on time preference provides an important theoretical context for our paper. Classical economic models of intertemporal choice postulate that individuals discount future events, relative to events today, in a time-consistent manner. That is, if future choice A is preferable to future choice B now, choice A will be preferable to choice B at all time periods. Putting this in a concrete example of a MOOC student, if she decides on Monday that she will watch the first MOOC lecture video on Thursday instead of joining friends at a bar, on Thursday her preference for the MOOC lecture over the bar will still hold.

Empirical studies from behavioral economics suggest people do not always conform to this theory and that preferences can change over time (Ariely & Wertenbroch, 2002; Bisin & Hyndman, 2014; DellaVigna, 2009; Frederick, Loewenstein, & O'Donoghue, 2002; Laibson, 1997; Loewenstein & Thaler, 1989; Loewenstein & Prelec, 1992; Thaler, 1981). These are called time-inconsistent preferences and typically result in a present bias. Using our example of a MOOC student, this means that her preferences for MOOC lecture watching compared to bar going might be different on Thursday than they were on Monday. The empirically demonstrated pattern of high initial MOOC registration combined with low course persistence suggests that many MOOC registrants may have time-inconsistent preferences.

Many people are aware of their present biased nature and value the opportunity to commit to a specific future behavior. These pledges are called *precommitment devices* because they attempt to impose self-control by restricting the future self. Our scheduling nudge can be considered a form of a precommitment device.² We hypothesize that the opportunity to precommit to watching a lecture at a specific day and time might provide MOOC registrants with an attractive and effective self-control mechanism that is consistent with their original intent to participate in the course. Several studies have provided empirical evidence that the opportunity for precommitment can change behaviors, such as employee effort (Kaur, Kremer, & Mullainathan, 2015), smoking (Gine, Karlan, & Zinman, 2010), and savings (Ashraf, Karlan, & Yin, 2006). One study from higher education (Ariely & Wertenbroch, 2002) has demonstrated a demand for, and a positive effect of, precommitment devices that aim to affect student effort and behaviors on out-of-class assignments.

Despite the existing empirical examples, the literature is in need of additional evidence on the efficacy of commitment devices for at least two reasons. First, results are inconsistent. Although several of the studies referenced above report positive effects, other studies (e.g., Bernheim, Meer, & Novarro, 2012; Bisin & Hyndman 2014; Bryan, Karlan, & Nelson, 2010; Burger, Charness, & Lynham, 2011; DellaVigna, 2009; Frederick et al., 2002) find little evidence that individuals routinely take up commitment devices or that they are effective at changing behavior. Second, it is likely that the effects of commitment are heterogeneous across individuals and settings; however, such heterogeneous effects have not been studied in great detail. In this study, information about the MOOC participants (time of registration, e-mail address, location, and early interaction with the course) allows us to examine how the effects of this treatment appear to vary by subject traits that are likely to be salient moderators of this treatment.

The only other test of a commitment device in an online educational setting of which we are aware is a smaller experiment conducted by Patterson (2014) in a MOOC. That experiment assigned treated students to a much more costly commitment: installing software on their computers that limits access to distracting websites (news, Facebook, etc.) to commit them to spend more time in the MOOC. The experiment finds large effects of the treatment, including an 11-percentage-point increase in course completion and an increase in course performance of more than a quarter of a standard deviation. However, there are substantial external validity limitations to the study. First, it randomizes only the small subset of course registrants (18%) who volunteered for the commitment device after being offered a financial incentive. Second, the commitment device is extremely intrusive as it requires the installation of third-party software that tracks and limits Internet activity. This

type of commitment device is unlikely to be highly scalable. Although the scheduling device tested in our study is relatively weak, it offers a more policy-relevant test of a nudge to increase persistence. The current study also enables a test of uptake and treatment effects across the full range of MOOC registrants, unlike Patterson.

In summary, our paper makes a contribution to the education and economics literatures by providing the first large-sample, causal evidence of the effect of asking students to schedule study time in an online course setting. The unique experimental setting of a MOOC allows for an examination of more fine-grained outcomes than is traditionally available in education research. Instead of focusing exclusively on GPA and course grades, we can also examine week-to-week course engagement and persistence. Given that the vast majority of MOOC research has focused on descriptive and correlational analyses (Veletsianos & Shepherdson, in press), we make an important contribution to the MOOC literature by testing a nudge intervention designed to improve student outcomes.

Experimental Design

Setting and Data

The course in this study, a science-based MOOC, was offered for the first time on the Coursera platform in 2013. The course was not overly technical, and there were no prerequisites. Students had two options for earning a certificate of accomplishment for the course: (a) satisfactorily completing weekly quizzes and weekly problem sets or (b) completing weekly quizzes and a final project. Both options required students to achieve an overall grade of at least 70% to earn the certificate.

The course was broken into eight topics, each lasting 1 week. Every week, the instructor released approximately a dozen video lectures, each approximately 10 to 20 minutes in length, which provided the bulk of the course content. Students could stream the video lectures directly through the Coursera website or download them for subsequent viewing offline. In addition to the lecture videos and the assignments, the course also had a discussion forum in which students could post questions or comments. Fellow students or the instructor could respond to these discussion threads. Although more students subsequently registered, we limited the experimental sample to the 18,043 students who registered at least 2 days before the course launched. This enabled us to implement our randomization and field the treatment contrast just prior to the course's official beginning. One of the challenges in conducting analyses using MOOC data is that Coursera does not collect any demographic information on course registrants, so we did not directly observe information such as age, gender, education level, or income. However, we gathered limited information on all registrants from the data that were available, such as the nature of the

registrants' e-mail address and the timing of their registration. In addition, we observed IP addresses for about 60% of registered students and used this information to determine whether they were domestic or international using the MaxMind IP address service, which geolocates IP addresses (http://www.maxmind.com/en/geolocation_landing). We also have demographic and motivational data from the subset of students (7.7%) who answered the instructor's pre-course survey. We provide a summary of the data for the full sample in Table 1.

Experimental Design and Implementation

We analyzed the effect of an intervention designed to increase student persistence by asking students to precommit to watching the first lecture of the week on a certain day and time of their choice. Students who registered for the course at least 2 days before it began were randomly assigned into treatment and control groups.³ Both groups received an e-mail from the course instructor with a link to a short online survey 2 days before the official start of the course. The e-mails were sent out automatically through the Qualtrics (2013) survey platform. The e-mail's from line was "[Course Title] Course Team" and signed by the instructor. The full treatment and control e-mails are included in Appendix A. The instructor was blind to the treatment status of each individual.

Students assigned to the treatment condition were sent an e-mail from the instructor that included a reminder that the course was starting and information about the requirements and expected time commitment for the 1st week (three videos and a quiz). Drawing from the literature on time management we cite above, the e-mail also stated, "Research finds that students who schedule their work time in advance get more out of their courses. Toward that end, please click on the link below to schedule when you intend to watch the first video entitled 'Week 1 introduction.'" The e-mail contained a link to an online survey that asked the student to commit to watching the first lecture of Week 1 on a day and time of their choice during the 1st week of the course. The survey asked students to respond to only two questions (both via radio buttons). One asked students to choose the day they planned to watch the first video, and the other asked students to indicate the hour (local time) in which they planned to watch the first video (a screenshot from the survey is available in Appendix A). The survey also noted that students should "make note of the day and time you selected on your calendar so you can get off to a good start in the course." Additionally, treatment students received a nearly identical second survey before the Week 2 videos were released asking them to precommit to a day and time to watch the first video of Week 2. Because each student received a unique link to the survey, we are able to connect student engagement with the scheduling survey to the course participation

TABLE 1
Descriptive Statistics

Variable	Mean
Watched first video of Week 1	0.468 (0.499)
Watched first video of Week 2	0.322 (0.467)
Number of unique lectures watched	19.025 (31.617)
Grade in course (0–100)	11.170 (27.284)
Earned a certificate	0.089 (0.285)
Assigned to treatment	0.500 (0.500)
Opened Week 1 scheduling survey: Assigned to T	0.125 (0.330)
Completed Week 1 scheduling survey: Assigned to T	0.101 (0.301)
Opened Week 2 scheduling survey: Assigned to T	0.065 (0.246)
Completed Week 2 scheduling survey: Assigned to T	0.049 (0.216)
Opened any scheduling survey: Assigned to T	0.159 (0.366)
Completed any scheduling survey: Assigned to T	0.127 (0.333)
Opened browser survey: Assigned to C	0.139 (0.346)
Answered browser survey: Assigned to C	0.135 (0.341)
@gmail e-mail address	0.561 (0.496)
.edu e-mail address	0.028 (0.164)
Completed the precourse survey	0.077 (0.266)
Registered early (≥ 1 week before course start)	0.569 (0.495)
Registered later (<1 week before course start)	0.224 (0.417)
Missing registration date ^a	0.207 (0.405)
Domestic student	0.170 (0.376)
International student	0.420 (0.494)
Missing country	0.409 (0.492)
<i>N</i>	18,043

Note. Standard deviations are shown in parentheses. Students' grades in the course are averages of their scores on quizzes and assignments. Students who had scores greater than 70 earned a certificate. Students' countries of residence were determined by geolocating their IP addresses. T = treatment; C = control.

a. We do not have good information on why 21% of students were missing the time of registration. This is not uncommon in MOOCs; in an analysis of 44 MOOCs across three universities, Evans, Baker, and Dee (2016) find similar rates of missing registration time.

data (e.g., we know if students watched the videos on the day and time that they committed to watching).

This intervention structure presents several possible ways of defining the treatment. The most general form of the treatment is having been sent two e-mails with links to scheduling surveys, one at the beginning of the 1st week and one at the beginning of the 2nd week. This definition conforms perfectly to treatment assignment and is presented in our intent-to-treat (ITT) estimates in the Results section. We cannot observe whether course registrants actually received and opened the e-mail; however, we do observe both whether registrants opened each of the scheduling surveys and

whether they submit the scheduling survey. Which survey action (opening or completing) we choose to call the treatment affects the rate of compliance to treatment and the treatment-on-the-treated (TOT) effects estimates. We examine results for defining the treatment in both ways.

We designed the control treatment with the goal of ensuring that communication with the instructor was consistent between the control and treatment groups, aside from the content of the treatment e-mail and survey. Students assigned to the control condition received an e-mail from the instructor at the beginning of the 1st week. This e-mail had the same content as the treatment condition 1st-week e-mail (salutation, reminder that the course was launching, and information about the workload for the 1st week) but replaced the scheduling text with a sentence asking students to report which web browser they used to access the course material (the text of the e-mail to control students is available in Appendix A). The e-mail contained a link to a survey asking them which web browser they used to access the course videos (one question with five radio buttons—a screenshot is in Appendix A). We intended this survey to be innocuous in that it did not convey any information about scheduling or timing. Students in the control condition did not receive a similar e-mail with survey link in the 2nd week of the course. They did, however, receive all other course communication: The course instructor sent two separate, nonsurvey e-mails to the whole class (treatment and control students) at the beginning of the 2nd week (once to announce that new course lectures were posted and once with additional clarifications and announcements).

One potential concern with this experimental setup is contamination. Students assigned to treatment and control conditions could communicate with each other via the course discussion forums. Students assigned to the control condition could learn of the scheduling device, and treatment students could learn that not all students were asked to schedule a time to watch. To assess the potential for and the seriousness of such a threat, we monitored the discussion forums for mentions of *schedule* over the course of the week after the first scheduling e-mail was sent. We find strong evidence that contamination is not a concern. Specifically, we found only one comment thread (out of hundreds of active threads) that mentioned the scheduling prompt. In this thread (which had only 308 total views), one student wondered about not having received or not noticing this prompt. This comment elicited no further discussion.

Blocking, Estimation, and Covariate Balance

The experimental sample contains the 18,043 students who had valid e-mail addresses and registered for the course at least 2 days before it began. The randomization process involved blocking on three variables: whether students completed the instructor’s precourse survey and whether their

e-mail was a gmail.com or .edu e-mail address. After blocking on these three variables, students were equally distributed between treatment ($n = 9,022$) and control ($n = 9,021$) groups. The response rate for the first survey is slightly higher for the control survey (13.9% opened the survey and 13.5% provided an answer) than the treatment survey (12.5% opened the survey and 10.1% selected a day and time to watch).

We employed the following regression model to estimate results.

$$y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{precourse}_i + \beta_3 \text{gmail}_i + \beta_4 \text{edu}_i + \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i. \quad (1)$$

We use y to denote our five outcomes: three measures of student engagement (watching the first lecture video of the 1st week, watching the first lecture video of the 2nd week, and the number of unique lectures the student watched), performance as measured by the final grade in the course on a scale of 0 to 100, and an indicator for earning the certificate of accomplishment, a measure that captures the intersection of student engagement and high achievement.⁴ We include the three blocking variables discussed above: completing the precourse survey and two types of e-mail addresses. Additionally we include pretreatment covariates. We use linear probability models to estimate the binary outcomes.

We have two sets of pretreatment covariates for all students included in the regression: registration time and country of location (obtained from IP addresses). We divide time of registration into three groups: at least 1 week before the classes started, less than a week before the class started, and missing registration time. We divide the country from which the student is accessing the course into domestic, international, or missing country. We use these covariates to assess the balance of treatment and control groups by regressing treatment status on this set of covariates and find none that are individually significant at the 10% level. Additionally, an F test cannot reject the null hypothesis that the covariates are jointly unrelated to treatment status, indicating that the groups appear to be balanced on the observable characteristics available in our data set. These covariate balance results are shown in Table 2.

Power

One of the advantages of conducting research in MOOCs is the large sample sizes, which generate enough power to detect even very small effects. As seen in our results below, we can identify statistical significance for effects of less than 1 percentage point in the probability of receiving a course certificate. As always, statistically significant results should not necessarily be interpreted as practically significant results. However, in the context of MOOCs, in which tens of thousands of students participate, small effects that might be of little concern in a traditional classroom have greater

TABLE 2
Predicting Treatment Status With Pretreatment Characteristics

Variable	Coefficient
Registered later (<1 week before course start)	-0.013 (0.009)
Missing registration date	-0.015 (0.012)
International student	-0.012 (0.011)
Missing country	-0.006 (0.012)
@gmail e-mail address ^a	0.001 (0.008)
.edu e-mail address ^a	-0.001 (0.023)
Completed the precourse survey ^a	-0.001 (0.014)
<i>N</i>	18043
Joint <i>F</i> test (<i>F</i> statistic)	0.64
Joint <i>F</i> test (<i>p</i> value)	0.7228
<i>R</i> ²	0.0002

Note. Robust standard errors are presented in parentheses. Registered later (at least 1 week before course start) and domestic students are the reference categories. Students' countries of residence were determined by geolocating their IP addresses.

a. Variable was used as a blocking variable in the randomization process.

consequence due to the number of students in one course. We believe the effects we observe are of practical significance, especially for the higher effect magnitudes observed in certain subgroups.

Limitations

As a randomized control trial, internal validity is extremely high. Attrition is not a concern as we observe all outcomes for the full analytic sample from the platform's administrative data. Experimental effects are unlikely as the participants were unaware they were part of an experiment. Spillover effects are also unlikely due to the lack of contact students have with each other. As mentioned previously, we monitored the discussion forum for mention of scheduling and found no evidence of contamination.

There are two primary limitations in this study: our inability to investigate standard subgroups and potential concerns about external validity. Because MOOCs capture very limited demographic data, standard subgroup analysis by gender, race, educational attainment, and socioeconomic status is not possible. Instead, we investigate subgroups of course behavior, such as registration time, in addition to the individual-level observable characteristics, such as domestic/international and type of e-mail address.

External validity limitations arise because we study a single course on one platform. We do not believe students meaningfully sort across different MOOC platforms, but they may across different courses. MOOC taking motivations differ across course discipline categories. Students in science MOOCs have fairly similar motivations to social science MOOC students, with approximately equal numbers being motivated by gaining knowledge for a degree (about

16%) or for simple curiosity (about 49%). However, social science students are more motivated by gaining skills for a job than science students. Humanity students are mostly motivated by curiosity or fun (Christensen et al., 2013).

Ho et al. (2014) examine the demographics of students in a small number of MIT and Harvard EdX courses. They identify trends of far fewer women than men (fewer than 20% in most cases) enrolling in computer science and physical sciences courses relative to life science and nonscience MOOCs, which have close to an equal gender balance. Importantly, most of the courses they studied are technical science courses.

However, even this concern of generalizability across MOOCs seems less important in this context. Given that the structure of the course we study is very similar to other MOOCs, and the fact that the subject of this MOOC (an accessible science topic) is similar to many of the courses described by Ho et al. (2014), we have no reason to believe that students who enroll in this course would be significantly different, in motivation or preparation, from the modal MOOC student. In terms of student motivations, our non-technical science course is likely similar to most social science or science MOOCs but likely has less direct job application than many others. Because we do not have demographic information on the respondents in our course, we cannot say definitely whether the population in our course is similar to those observed elsewhere, but we have little reason to believe it is an outlier.

Results

Treatment Uptake

Previous literature suggests uptake of commitment devices can be low. We examine treatment uptake in our experiment via our first-stage estimates in Table 3. As noted above, we combined the 1st- and 2nd-week treatments to get a measure of whether a student opened or completed either scheduling survey and used this as our measure of treatment. Although the intended treatment is scheduling a time to watch the lecture video, it is also possible that simply opening the survey to view the scheduling options constitutes a treatment. Table 3 provides treatment uptake estimates for each definition of treatment (opening and completing the survey) for three models: a bivariate regression of the dependent variable on only treatment assignment, the same regression with the inclusion of blocking covariates, and a regression including all available covariates. Results are consistent across these three models. As control students did not have access to these scheduling surveys, the control mean is 0. We find that nearly 16% of treatment students opened at least one of the scheduling surveys and 12.7% completed a survey and committed to watching the lecture video on a specific date and time.⁵ Relative to the 7.7% who completed the

TABLE 3

First Stage Estimates: The Determinants of Treatment Uptake

Independent variable	Opened any scheduling survey			Completed any scheduling survey		
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned to treatment	.159*** (.004)	.159*** (.004)	.159*** (.004)	.127*** (.004)	.127*** (.003)	.127*** (.003)
@gmail e-mail address		-.007 [†] (.004)	-.005 (.004)		-.005 (.004)	-.003 (.004)
.edu e-mail address		-.002 (.012)	-.011 (.012)		-.005 (.011)	-.014 (.011)
Completed precourse survey		.152*** (.010)	.141*** (.010)		.138*** (.010)	.128*** (.010)
Registered later (<1 week before course start)			.023*** (.005)			.021*** (.005)
Missing registration date			.061*** (.005)			.046*** (.004)
International student			-.028*** (.006)			-.032*** (.006)
Missing country			-.106*** (.006)			-.094*** (.006)
R^2	.086	.109	.128	.068	.091	.108

Note. The sample size is 18,043. Regression coefficients are reported with robust standard errors in parentheses. Registered early (at least 1 week before course start) and domestic students are the reference categories. Students' countries of residence were determined by geolocating their IP addresses.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

precourse survey, treatment uptake is quite high, likely because of the ease of the survey. However, the overall modest uptake of this treatment is itself a substantive finding and consistent with recent precommitment studies. We have suggestive evidence that the students who completed the scheduling survey did intend to follow through; 85% of students who filled in the first survey did watch the first video, and students generally selected reasonable times to engage in a leisure activity (evenings).

Focusing on the group of students who completed the first scheduling survey, we examine when they committed to watching the first video of the 1st week of the course and whether they complied with this commitment. Figure 1 presents a density plot of the time students indicated that they would watch the first video after it was released (time as a continuous variable on the x -axis, the number of students who committed to watching at that time on the y -axis). We see consistent temporal patterns: Scheduled watching peaks the day after the video was released and then falls over time. There are distinct peaks in the evening of each day (we have labeled 9:00 p.m. each day). Among students who completed the survey and actually watched the first video, Figure 2 depicts the day and time at which students watched the first video relative to when they committed to watching it. The left histogram confirms that the majority of students watch the video on the day they scheduled with a significant number watching either a day early or late. The right panel shows

that for students who watched on the day they scheduled to watch, the majority watched it within an hour of the time they scheduled. Students who commit to watching the video on a specific day and time are mostly adhering to their schedule.

Main Results

We now turn to whether random assignment to the treatment—being encouraged to precommit to a specific time to watch the first video—affected lecture watching and measures of persistence and performance in the course. We first present ITT estimates in Table 4. These are results from linear regressions of each of these five outcomes on treatment assignment, the blocking variables, and four additional covariates.⁶ The estimates in this table show that there were no effects of the scheduling device on near-term outcomes and weakly significant negative effects for later outcomes. Assignment to treatment did not have a statistically significant effect on the proximate outcomes of watching the first video of the 1st or 2nd week of the class, but there were marginally significant negative effects on the longer-term persistence and performance measures of total number of lectures watched, course grade, and certificate completion. Relative to control students, students assigned to treatment watched three quarters of a video less, had over-a-half-point-lower course grade, and were nearly 10% less likely to earn

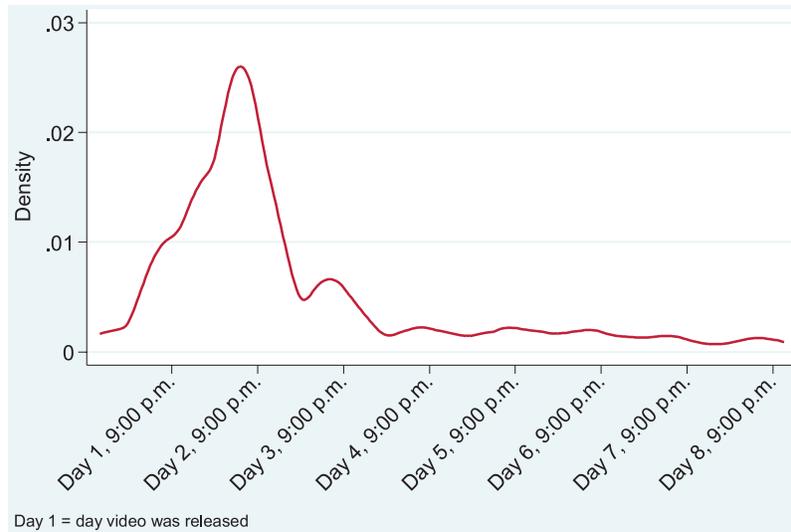


FIGURE 1. *When students committed to watch lectures.*

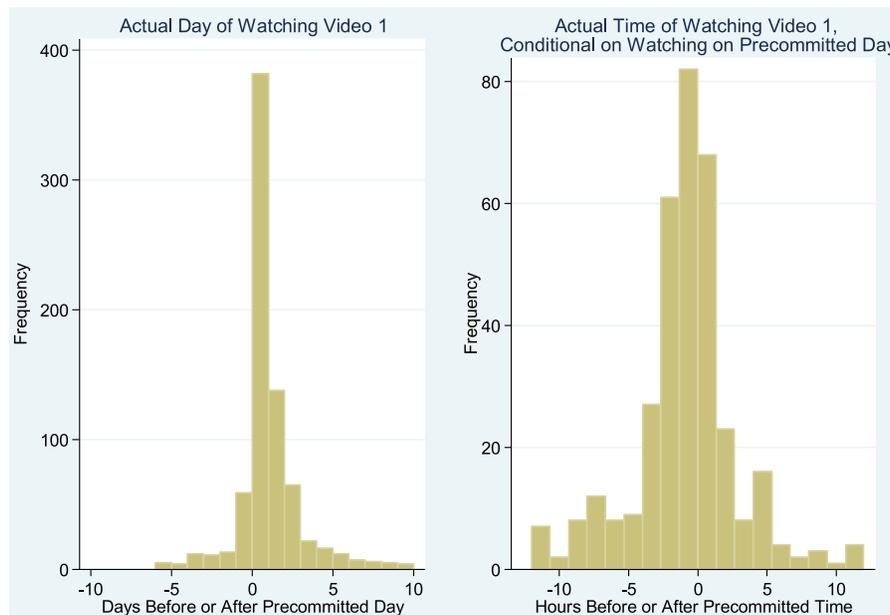


FIGURE 2. *Day and time of watching Video 1 relative to precommitment day and time.*

a certificate (from a baseline of 9.3% of control students earning a certificate). This finding that the results were significant only for the more distal outcomes is interesting and might reflect the dynamic nature of engagement; as course participation is a recursive process, this early treatment could have an effect that grew over time as students continued to interact with the course.

These ITT effects are likely diluted by the majority of registrants in the treatment group who did not open or respond to the scheduling surveys. Although it is possible to conceive of the treatment as simply reading the e-mail (which we cannot

observe), we believe it more likely that the treatment was far more salient for those that opened and scheduled the survey. Therefore, it is also important to consider the effect on those who actually took up the treatment, the implied TOT estimates. As we discussed above, there are several ways to define uptake of treatment. At one extreme, we could conceive of treatment as reading the e-mail (being told that one should schedule a time to watch the first lecture). If we take this as the treatment and assume that every student who was sent the e-mail read it, the TOT estimates of the effect of the treatment are equal to the ITT estimates. This serves as the

TABLE 4

Intent-to-Treat Estimates: Determinants of Course Persistence and Performance

Independent variable	Outcome				
	Watched first video of Week 1	Watched first video of Week 2	Number of unique lectures watched	Grade in course (0–100)	Earned a certificate
Assigned to treatment	−0.010 (0.006)	−0.003 (0.006)	−0.733 [†] (0.419)	−0.642 [†] (0.381)	−0.008 [†] (0.004)
@gmail e-mail address	−0.048*** (0.006)	−0.059*** (0.006)	−4.696*** (0.438)	−4.006*** (0.402)	−0.039*** (0.004)
.edu e-mail address	−0.091*** (0.020)	−0.115*** (0.019)	−8.322*** (1.224)	−4.851*** (1.124)	−0.042*** (0.012)
Completed precourse survey	0.198*** (0.011)	0.190*** (0.013)	12.100*** (0.939)	14.670*** (0.980)	0.118*** (0.011)
Registered later (<1 week before course start)	0.006 (0.008)	−0.005 (0.008)	−1.671** (0.605)	1.142* (0.560)	0.009 (0.006)
Missing registration date	0.372*** (0.008)	0.145*** (0.006)	3.874*** (0.198)	1.393*** (0.170)	−0.005*** (0.001)
International student	−0.080*** (0.010)	−0.069*** (0.011)	−2.933*** (0.778)	−0.294 (0.705)	0.003 (0.008)
Missing country	−0.693*** (0.008)	−0.518*** (0.009)	−31.568*** (0.671)	−16.918*** (0.601)	−0.136*** (0.006)
Mean of outcome for control group	0.474	0.322	19.37	11.49	0.093
R^2	0.283	0.214	0.209	0.122	0.085

Note. The sample size is 18,043. Regression coefficients are reported with robust standard in parentheses. Registered early (at least 1 week before course start) and domestic students are the reference categories. Students' grades in the course are averages of their scores on quizzes and assignments. Students who had scores greater than 70 earned a certificate. Students' countries of residence were determined by geolocating their IP addresses.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

lower bound of the TOT estimates. At the other extreme, we could consider actually completing the survey as the treatment and use the proportion of students who completed the survey to compute the TOT estimates. This serves as an upper bound for our TOT estimates. In Table 5, we present estimates using opening the survey as the definition of treatment, which falls somewhere between the upper and lower bounds. Using completing any scheduling survey as the treatment inflates the TOT estimates slightly but maintains significance. Estimates are available in Table 6.

Using this definition of treatment, the full-sample TOT estimate implies that the causal effect of taking up the treatment on the certificate outcome was negative 4.8 percentage points, which is significant at the 10% level. As a point of reference, the probability of earning a certificate in the control group was 9.3%. Relative to this baseline, treatment uptake decreased this measure of long-term persistence and engagement by 52%. Similarly, the corresponding TOT estimates imply that taking up the treatment reduced course grades by 4.0 percentile points relative to a control group mean of 11.5. Finally, it reduced the number of lectures watched by 4.6 relative to a control group mean of 19. Both the signs and magnitudes of these experimental effects are somewhat surprising. These patterns suggest that the

possible heterogeneity in these treatment effects might be informative.

Treatment Heterogeneity

As noted earlier, Coursera collected very little data consistently for all the students who registered for their classes. For this reason, we did not have many groups for whom we could assess treatment heterogeneity in our full analytical sample. Table 5 presents estimates of treatment heterogeneity for the available variables: students' type of e-mail address, time of registration, country from which they accessed the course, and whether they completed a precourse survey offered by the instructor. The first column of results in Table 5 provides first-stage estimates of opening the scheduling survey for each subgroup, and the subsequent columns provide TOT estimates for that subgroup across all five outcomes. All estimates used models that included the full set of covariates. When considering treatment heterogeneity, it is helpful to consider the local average treatment effect introduced by Imbens and Angrist (1994). The causal effect of the treatment is identified for the subpopulation of "compliers" who take up the treatment. As demonstrated in the first column in Table 5,

TABLE 5

Reduced-Form and First-Stage Estimates for Persistence Outcomes by Sample Traits

Independent variable	First stage	Treatment on the treated					Sample Size
	Opened any scheduling survey	Watched first video of Week 1	Watched first video of Week 2	Number of unique lectures watched	Grade in course (0–100)	Earned a certificate	
Full sample	0.159*** (0.004)	−0.062 (0.040)	−0.017 (0.039)	−4.600 [†] (2.648)	−4.029 [†] (2.408)	−0.048 [†] (0.026)	18,043
@gmail e-mail	0.149*** (0.005)	−0.063 (0.057)	0.027 (0.054)	−4.223 (3.576)	−3.86 (3.122)	−0.037 (0.033)	10,128
.edu e-mail	0.166*** (0.023)	−0.461 [†] (0.249)	−0.487* (0.238)	−38.036* (15.190)	−29.593* (13.781)	−0.246 [†] (0.141)	498
Other e-mail	0.173*** (0.006)	−0.027 (0.056)	−0.030 (0.057)	−2.181 (4.033)	−1.838 (3.823)	−0.042 (0.041)	7,425
Precourse survey	0.436*** (0.019)	−0.041 (0.047)	−0.036 (0.055)	−7.219 [†] (4.181)	−3.826 (4.348)	−0.057 (0.049)	1,386
No precourse survey	0.136*** (0.004)	−0.070 (0.049)	−0.015 (0.047)	−4.150 (3.165)	−4.393 (2.819)	−0.048 (0.030)	16,657
Early registrant	0.146*** (0.005)	−0.053 (0.054)	−0.002 (0.056)	−2.357 (4.194)	−1.204 (3.758)	−0.027 (0.040)	10,261
Later registrant	0.209*** (0.009)	−0.059 (0.066)	−0.047 (0.069)	−10.836* (4.955)	−11.584* (4.687)	−0.112* (0.050)	4,042
Missing registration date	0.143*** (0.008)	−0.078 (0.111)	−0.001 (0.083)	−0.159 (1.909)	0.424 (1.423)	−0.004 (0.004)	3,740
Domestic student	0.253*** (0.011)	−0.089 (0.064)	−0.093 (0.070)	−8.276 (5.201)	−7.018 (4.715)	−0.073 (0.051)	3,076
International student	0.197*** (0.006)	−0.038 (0.056)	0.022 (0.057)	−3.949 (4.239)	−3.925 (3.875)	−0.050 (0.042)	7,585
Missing country	0.081*** (0.004)	−0.074 (0.100)	−0.008 (0.075)	−0.505 (1.786)	−0.318 (1.355)	−0.010 [†] (0.006)	7,382

Note. Treatment-on-the-treated results report the treatment variable’s coefficient and robust standard errors of two-stage least squares estimates of for each outcome and subgroup. Both first-stage and reduced-form results are from regressions including all available covariates. First-stage estimates employ opening any scheduling survey as the treatment. Precourse survey is a dummy with 1 = the student completed the precourse survey sent out by the instructor. Early registrants are students who registered 1 week or more before the course started. Later registrants are those who registered less than 1 week before the course started. Students’ grades in the course are averages of their scores on quizzes and assignments. Students who had scores greater than 70 earned a certificate. Students’ countries of residence were determined by geolocating their IP addresses.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

take-up is inconsistent across subgroups, implying that the ITT estimates for each group will be inflated by different amounts.

We observe interesting patterns of treatment heterogeneity, although for no group do we find a positive effect of the scheduling device. Evans et al. (2016) found that the best predictor of course persistence is completion of the instructor’s precourse survey. These students are typically the most engaged throughout the course, so it is no surprise that they have the highest uptake of treatment across all subgroups we analyzed (44% opened at least one scheduling survey). However, the treatment appears to have had minimal effect on these students: The point estimates are consistently negative, but only one (number of lectures watched) is marginally significant. The soft commitment

device tested in our experiment did not affect this group of highly motivated students.

The negative results observed for the full sample appear to be driven by two groups: those who registered within a week of the course beginning (later registrants) and those with .edu e-mail addresses. There are large and statistically significant negative outcomes across both of these subgroups. Later registrants who took up the treatment watched on average 11 fewer videos, scored nearly 12 points lower on their final grade, and were 11 percentage points less likely to receive a certificate relative to later registrants who were not offered the scheduling treatment. It is important to remember that these “later registrant” students still registered before the course began, so these results were not driven by their having to catch up on missed work.

TABLE 6

Reduced-Form and First-Stage Estimates for Persistence Outcomes by Sample Traits

Independent variable	First stage		Treatment on the treated				Sample size
	Completed any scheduling survey	Watched first video of Week 1	Watched first video of Week 2	Number of unique lectures watched	Grade in course (0–100)	Earned a certificate	
Full sample	0.127*** (0.003)	−0.078 (0.050)	−0.021 (0.049)	−5.778 [†] (3.328)	−5.061 [†] (3.027)	−0.06 [†] (0.032)	18,043
@gmail e-mail	0.119*** (0.004)	−0.078 (0.071)	0.034 (0.067)	−5.282 (4.475)	−4.828 (3.908)	−0.046 (0.041)	10,128
.edu e-mail	0.124*** (0.020)	−0.616 [†] (0.341)	−0.651* (0.326)	−50.859* (20.884)	−39.57* (18.828)	−0.329 [†] (0.192)	498
Other e-mail	0.137*** (0.006)	−0.034 (0.071)	−0.037 (0.072)	−2.742 (5.073)	−2.311 (4.809)	−0.053 (0.052)	7,425
Precourse survey	0.377*** (0.018)	−0.047 (0.055)	−0.042 (0.064)	−8.355 [†] (4.847)	−4.428 (5.035)	−0.066 (0.057)	1,386
No precourse survey	0.106*** (0.003)	−0.09 (0.063)	−0.019 (0.060)	−5.338 (4.072)	−5.65 (3.629)	−0.062 (0.038)	16,657
Early registrant	0.117*** (0.004)	−0.067 (0.067)	−0.003 (0.070)	−2.942 (5.237)	−1.504 (4.693)	−0.034 (0.050)	10,261
Later registrant	0.173*** (0.008)	−0.071 (0.080)	−0.057 (0.083)	−13.098* (5.999)	−14.001* (5.682)	−0.135* (0.061)	4,042
Missing registration date	0.105*** (0.007)	−0.107 (0.151)	−0.002 (0.113)	−0.217 (2.601)	0.578 (1.937)	−0.005 (0.005)	3,740
Domestic student	0.219*** (0.011)	−0.102 (0.074)	−0.107 (0.081)	−9.546 (6.002)	−8.095 (5.444)	−0.084 (0.058)	3,076
International student	0.156*** (0.006)	−0.048 (0.071)	0.028 (0.073)	−4.99 (5.357)	−4.96 (4.899)	−0.063 (0.053)	7,585
Missing country	0.058*** (0.004)	−0.103 (0.140)	−0.012 (0.105)	−0.706 (2.497)	−0.444 (1.894)	−0.014 [†] (0.008)	7,382

Note. Treatment-on-the-treated results report the treatment variable's coefficient and robust standard errors of two-stage least squares estimates of for each outcome and subgroup. Both first-stage and reduced-form results are from regressions including all available covariates. First-stage estimates employ completing any scheduling survey as the treatment. Precourse survey is a dummy with 1 = the student completed the precourse survey sent out by the instructor. Early registrants are students who registered 1 week or more before the course started. Later registrants are those who registered less than 1 week before the course started. Students' grades in the course are averages of their scores on quizzes and assignments. Students who had scores greater than 70 earned a certificate. Students' countries of residence were determined by geolocating their IP addresses.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

There were even stronger negative results for students with .edu e-mail addresses, although they composed only a very small subset of the full sample (498 students, 2.8% of the full sample). The treatment caused significant negative outcomes for these students across all five outcomes we observed. The probability that treated students with .edu e-mail addresses watched the first lecture videos of Week 1 and Week 2 were 46 to 49 percentage points lower than those students with .edu e-mail addresses not exposed to the treatment. They also watched many fewer lecture videos in total (38 fewer), had much lower grades (nearly 30 points lower), and were much less likely to earn a certificate (nearly 25 percentage points less likely).

We have limited information about these later registrants and .edu students that could help us isolate the potential

mechanism for the negative treatment effect on these groups. The .edu students might have been students enrolled in traditional higher education institutions, retired or active faculty, or other employees affiliated with a college or university. Very few had clearly alumni accounts (e.g., alumni.edu, post.edu), suggesting that alumni did not constitute a substantial portion of this group. To the extent that .edu students consisted of traditional college students, their primary concern may have been the completion of their credit-bearing courses. They might have been using this MOOC simply as an ancillary support for their other studies. We hypothesize that these students were less motivated to complete the course, and therefore, the commitment device might have signaled greater seriousness than they were expecting and thus turned them away.

The same theory may apply to later registrants who may also have been checking out the course as part of an exploration of the MOOC universe and chose this course simply because it was about to launch. Coursera often advertises courses that will be launching soon on its website, so later registrants may have seen this course advertised on the front page and decided to enroll without high levels of motivation to complete it. As a casual participant with little ambition to complete the full course, they may have been turned off by the suggestion to commit to schedule and watch lecture videos. We discuss this possibility more in the next section.

Discussion

The consensus in the online education literature is that time management skills are crucial for success in online, asynchronous learning environments, such as MOOCs. We attempted to intervene in such an environment to experimentally test whether giving students the opportunity to schedule a time to watch lecture videos in an online class affects course engagement, persistence, performance, and completion. To be clear, we did not assign time management skills. Instead, the intervention was an attempt to help students develop time management skills. We found weakly significant, modestly negative effects of asking students to schedule time to commit to the course in advance on longer-term measures of engagement. Those effects appeared concentrated in two subgroups: those with .edu e-mail addresses and those who registered for the course in the week before it started.

The only other test of a commitment device in a MOOC of which we are aware found large, positive effects on course completion of committing to block distracting websites when engaged with the MOOC online (Patterson, 2014). The contrast in findings from our null or negative results suggests that both the strength of commitment devices and student motivation are important. Although neither commitment device attaches direct financial penalties for violating the commitment, Patterson's (2014) device is strong and intrusive (monitoring and blocking web browsing) relative to the soft nudge to schedule time to work on the course studied in the current paper. Although a small subset of students may find such a strong commitment useful, it is important to discover nudges that could be more widely adopted. Patterson's study also randomly assigned only students who had expressed interest in taking part in a study on time management, whereas our study randomly assigned all students enrolled in a MOOC. Thus, student motivation and expectations might interact with theoretically sound interventions in important and unintended ways.

Our study has high internal validity due to the randomized controlled trial design; therefore, we are confident the observed results were causally driven by the suggestion to schedule lecture watching. Yet the empirical results are

surprising. Our expectation, grounded in theory and prior literature, was that students would benefit from a suggestion to schedule as it would impose a greater degree of structure in an otherwise fairly structureless learning environment (Elvers et al., 2003). A null finding could be explained more easily, as the suggestion to schedule to watch two lecture videos might not have an observable effect on long-term course outcomes. Negative findings suggest another mechanism was at play.

The experimental findings imply one or more in a set of explanations is at work. First, it is possible that students in this MOOC did not have time-inconsistent preferences such that a commitment device is useful. Given the ubiquity of time-inconsistent preferences across many contexts, including in education, we find this unlikely. Second, students might not be aware of their time-inconsistent preferences. Commitment devices should help only students who are sophisticated enough to anticipate changing preferences over time and therefore find committing to a future activity valuable. It is possible this explains the low uptake of the scheduling survey; students do not anticipate needing help scheduling. Both of these explanations could justify a null finding, but neither of these explanations rationalizes a negative effect of the scheduling survey.

Third, it is possible that the commitment does not have enough of a penalty to motivate follow-through. The penalty for violating the commitment to watch the video at the scheduled time is purely psychological; there was no academic or economic penalty associated with violating the set schedule. Students may have worried that the course instructor would think less of them for missing the deadline, but without any face-to-face interaction and given the widely known enormous size of MOOCs, it seems unlikely that students believed the instructor was tracking which students did and did not meet their self-imposed schedule commitments. Either way, like the previous two, this explanation again suggests a null finding, not negative results.

These negative effects on longer-term outcomes raise a number of interesting questions and directions for future research. They suggest that the role of time inconsistency in explaining the poor persistence of MOOC students should be considered uncertain. The heterogeneity of our effects—negative effects among small, defined segments of MOOC registrants (e.g., late registrants)—are also a provocative and potentially generative finding. Below we provide an outline of a research agenda that would serve to replicate and extend the experiment in this paper. We organize this section by possible explanation for the surprising negative findings and then provide a viable way to test whether that explanation is a likely cause of the observed results.

Before considering plausible mechanisms underlying for our findings, it is worth noting that the take-up rate of the opportunity to precommit was low; just under 13% of students named a day and time to watch the first lecture video.⁷

This number is consistent with Gine et al. (2010), who found 11% of those offered an opportunity to take up a financial precommitment to stop smoking, but our take-up rates were substantially lower than Bisin and Hyndman (2014), who found the demand for setting deadlines to be between 30% and 60%. However, Bisin and Hyndman found more robust demand for commitment for more complex tasks over a longer time horizon. Ours was a simple, proximate task, so low demand for commitment broadly aligns with previous findings.

Motivation for Taking the Course

Given the nature of the MOOC in this study—a science course not directly related to skills for a job—it is possible that the majority of students enrolled for fun (indeed, such a conjecture is supported by responses to the precourse survey—the vast majority of respondents cited taking the course because it seemed fun and interesting). Negative results could arise, particularly for lightly motivated students, if the scheduling nudge signaled a seriousness of purpose that did not match students' view of the course. Perhaps watching subsequent videos felt like a chore instead of a choice. Given this, the time management strategies promoted by our intervention might be much more effective for students enrolled in a course for academic or professional purposes. We propose replicating the experiment in two different types of online courses: a MOOC that has a distinctly academic or professional purpose and a traditional online course offered by a university to its enrolled students for credit. Observing negative findings in these contexts would rule out motivation as an explanation for the negative effect of the scheduling intervention.

Mismatch of Expectations

One of the advantages of a MOOC is it allows a flexible schedule due to the asynchronous nature of the course. We believe it is possible that the professor's suggestion to schedule turned off a subset of students for whom that flexibility is seen as a primary advantage of online learning. A similar explanation is that the MOOC offers flexibility to pick and choose to engage with only certain topics throughout the entire course. Perhaps the suggestion to schedule to watch the first 2 weeks conflicted with the goal of watching only a few of the later weeks of the course. Testing both of these expectations as primary drivers of the negative result is possible by using a more robust precourse survey that asks students about their expectations for engaging with the complete course and whether the flexibility to not schedule in advance is a desirable feature of the MOOC. By blocking on high and low responses to these questions on the precourse survey, we could ensure equal distribution of different types of students into treatment and control and then evaluate if effects are stronger for those subgroups.

Guilt or Shame for Not Following Through on the Schedule Commitment

Although the majority of people who scheduled did actually watch the video close to their scheduled time, it is plausible the results are driven by a subset of people who did not follow through on their schedule. The students who missed their scheduled time may have felt guilt over having missed the schedule, or they might have feared some external shame from the instructor if they believe the instructor is monitoring their follow-through on the scheduling commitment. There is a variety of ways this feeling could be tested in a replication. First, we could minimize the feeling of guilt and shame for those who missed the schedule by informing them that many other people missed the schedule as well and by telling them that missing the schedule is not that important and that there is still ample time to watch the lecture. Alternatively, we could increase the feeling of shame by acknowledging that the professor has noted the student missed the schedule. We could also move the intervention closer to a hard precommitment device by establishing an academic penalty for not following through on the commitment, such as losing points on the next quiz, which would presumably increase the negative feelings associated with not following through on the schedule.

Opportunity to Schedule Felt Disrespectful

A subset of students may have felt that their time management skills were already well developed and the suggestion to schedule may have felt disrespectful to them. Feeling disrespected may demotivate them from continuing in the course, resulting in negative long-term outcomes. This mechanism could be tested by framing the scheduling prompt in different ways. The e-mail from the professor in one treatment arm could soften the prompt to say the course is offering this scheduling prompt as a tool and students are welcome to make use of it if they would like. In another treatment arm, the professor could more strongly suggest that students need this tool to succeed in the course.

Highlighting a Deficit

Research demonstrates that competing time pressures is a primary reason why students drop out of a MOOC (Nawrot & Doucet, 2014). Perhaps the negative results are driven by a subset of students who know they have trouble with time management and the suggestion to schedule is highlighting that deficit and reinforcing it, resulting in negative academic outcomes. We propose to test this by asking their perceived time management capabilities in a precourse survey and observing the results for this subgroup.

Poor Randomization

It is always possible that randomization resulted in an unbalanced group and that the negative results are driven by more students less likely to succeed in the treatment condition. We find this implausible given the large sample size, balance on observables, and our blocking strategy to ensure that the most motivated students (those who completed the precourse survey) are equally distributed between treatment and control. This explanation will be further tested through any replication with a new randomly assigned treatment condition.

In summary, we are puzzled by the results and see many fruitful avenues for future research. The negative results cannot easily be explained by economic or education theory on time preferences or time management. It is also somewhat surprising that the intervention negatively affects more distal outcomes, such as total lectures watched and certificate completion, without concomitant negative results for the outcomes immediately salient to the intervention (i.e., watching the first lecture videos of Weeks 1 and 2).

Although it is possible these results hold only for MOOC participants and do not widely apply to all types of online

education settings, we caution all online course designers and instructors to consider the unintended consequences of interventions designed to improve time management skills.

Conclusion

This study presents results from a randomly assigned scheduling nudge using high-frequency data in a novel and policy-relevant educational setting. The scalability of the intervention studied enhances its policy relevancy; however, the unexpected negative findings for the .edu and late-registrant subgroups underscore the ways in which ostensibly helpful interventions designed to promote student success can have unintended consequences. More generally, these results suggest that the design of behavioral nudges should be sensitive to the possible complications in how the offer of a nudge is interpreted by different individuals. The scheduling intervention may interact with student motivations, suggesting the intervention may be more successful in online contexts in which students are highly motivated by course completion and earning credit, which is not the case for most MOOC students. The additional research outlined in the Discussion section will further explore the mechanisms at work.

Appendix

*****TREATMENT*****

From: COURSE NAME Course Team

Subject: Week 1 videos

Reply to: noreply@coursera.org

Dear %NAME%,

I hope you, like me, are looking forward to the formal launch of the course. This is a reminder that the Week 1 lecture videos for the course will be available starting Sunday night, DATE, Pacific Time.

Research finds that students who schedule their work time in advance get more out of their courses. Toward that end, please click on the link below to schedule when you will watch the first video entitled "Week 1 introduction:"

{LINK}

The other required course activities for Week 1 include three additional lecture videos and the weekly review quiz. I estimate that this will take approximately two hours of work. For those of you planning to pursue the quantitative approach for the course, note that there will be no problem set for the first week.

I am looking forward to your participation!

Regards,

INSTRUCTOR NAME

FIGURE A1. *E-mail from instructor (treatment group).*

*****CONTROL*****

From: COURSE NAME Course Team
Subject: Week 1 videos
Reply to: noreply@coursera.org

Dear %NAME%,

I hope you, like me, are looking forward to the formal launch of the course. This is a reminder that the Week 1 lecture videos for the course will be available starting Sunday night, DATE, Pacific Time.

Please help us make your experience taking online classes better by clicking on the following link and answering the survey question about which browser you will use to watch the first lecture video:

{link}

The other required course activities for Week 1 include four lecture videos and the weekly review quiz. I estimate that this will take approximately two hours of work. For those of you planning to pursue the quantitative approach to the course, note that there will be no problem set for the first week.

I am looking forward to your participation!

Regards,
INSTRUCTOR

FIGURE A2. E-mail from instructor (control group).

I will start watching the "Week 1 introduction" video on:

<input type="radio"/> Sunday, April 7	<input type="radio"/> Thursday, April 11
<input type="radio"/> Monday, April 8	<input type="radio"/> Friday, April 12
<input type="radio"/> Tuesday, April 9	<input type="radio"/> Saturday, April 13
<input type="radio"/> Wednesday, April 10	<input type="radio"/> Sunday, April 14

at (local time):

<input type="radio"/> Midnight	<input type="radio"/> Noon
<input type="radio"/> 1:00 a.m.	<input type="radio"/> 1:00 p.m.
<input type="radio"/> 2:00 a.m.	<input type="radio"/> 2:00 p.m.
<input type="radio"/> 3:00 a.m.	<input type="radio"/> 3:00 p.m.
<input type="radio"/> 4:00 a.m.	<input type="radio"/> 4:00 p.m.
<input type="radio"/> 5:00 a.m.	<input type="radio"/> 5:00 p.m.
<input type="radio"/> 6:00 a.m.	<input type="radio"/> 6:00 p.m.
<input type="radio"/> 7:00 a.m.	<input type="radio"/> 7:00 p.m.
<input type="radio"/> 8:00 a.m.	<input type="radio"/> 8:00 p.m.
<input type="radio"/> 9:00 a.m.	<input type="radio"/> 9:00 p.m.
<input type="radio"/> 10:00 a.m.	<input type="radio"/> 10:00 p.m.
<input type="radio"/> 11:00 a.m.	<input type="radio"/> 11:00 p.m.

Please **make note of the date and time you selected on your calendar** so you can get off to a good start in the course

Press "SUBMIT!" below so that your answers are recorded

FIGURE A3. Scheduling survey (treatment group).

I will use:

- Internet Explorer
- Firefox
- Chrome
- Safari
- Opera
- Other

Press "SUBMIT!" below so that your answer is recorded.

FIGURE A4. Browser survey (control group).

Authors' Note

Rachel Baker and Brent Evans received support from Institute of Education Science Grant No. R305B090016 for this work. We thank the instructor of the course used in this experiment as well as the Stanford Lytics Lab for data access. Several people at the Center for Education Policy Analysis at the Stanford Graduate School of Education also provided insightful comments on this research.

Notes

1. The massive open online course (MOOC) in which our experiment is situated has persistence patterns that are typical. Forty-eight

percent of registrants watched the first lecture, and 9% earned the course certificate.

2. Bryan, Karlan, and Nelson (2010) extend the modeling of precommitments by making a compelling distinction between “hard” and “soft” commitments. They define hard commitments as those imposing an economic penalty for violating the commitment and soft commitments as ones in which the consequences are predominately psychological. The commitment that was randomly offered in this experiment—a request to schedule a day and time to watch the first lecture video of an online course—is a soft commitment. There is no economic penalty for violating the commitment to watch on the committed day and time, but there is a potential psychological penalty as students may not want to perceive of themselves as failing to follow through or as procrastinating on their preconceived goals. The failure to follow through on a scheduling commitment may create a cognitive dissonance that individuals would rather avoid.

3. The institutional review board approved this research by relying on the general Coursera consent language that all registrants must agree to be using the service in which they specifically agree to participate in research related to improving education.

4. Final grades in the course were an average of either students’ weekly review quiz scores (eight quizzes) and weekly problem set scores (seven problem sets) or their weekly review quiz scores and a final project (graded credit/nocredit). The lowest quiz and problem set scores were dropped for each student. Students who earned an average score of 70% on weekly quizzes and problem sets (or received an average score of 70% on quizzes and submitted a creative project) received a certificate of accomplishment.

5. Twelve-and-a-half percent of students in the treatment condition opened the first scheduling survey, and 80% of them (10.1% of all treatment students) selected a day and time to watch the first video of the 1st week. Six-and-a-half percent of students opened the second scheduling survey, and 76% of them (4.9%) selected a day and time to watch the first video of the 2nd week.

6. For ease of interpretation, we present estimates from linear probability models. We also examined the estimated marginal effects from probit models for the binary outcomes and negative-binomial models for the count-data outcomes and find substantively similar results.

7. Since the opportunity to take up the commitment in our study was delivered via e-mail, we cannot assess whether all participants in the treatment group actually read the e-mail and were made aware of the scheduling commitment opportunity, which could downwardly bias our measure of demand for commitment.

References

- Andrews, R., Li, J., & Lovenheim, M. F. (2014). Heterogeneous paths through college: Detailed patterns and relationships with graduation and earnings. *Economics of Education Review*, *42*, 93–108.
- Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*, 219–224.
- Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*, *121*, 635–672.
- Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary education. In G. Duncan & R. Murnane (Eds.), *Whither opportunity?* (pp. 117–132). New York, NY: Russell Sage Foundation.
- Banerjee, A., & Dufló, E. (2014). (Dis)Organization and success in an economics MOOC. *American Economic Review: Papers and Proceedings*, *104*, 514–518.
- Bernheim, B. D., Meer, J., & Novarro, N. K. (2012). *Do consumers exploit precommitment opportunities? Evidence from natural experiments involving liquor consumption* (NBER Working Paper No. 17762). Retrieved from <http://www.nber.org/papers/w17762>
- Bisin, A., & Hyndman, K. (2014). *Present-bias, procrastination and deadlines in a field experiment* (NBER Working Paper No. 19874). Cambridge, MA: National Bureau of Economic Research.
- Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, *83*, 405–410.
- Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment devices. *Annual Review of Economics*, *2*, 671–698.
- Burger, N., Charness, G., & Lynham, J. (2011). Field and online experiments on self-control. *Journal of Economic Behavior and Organization*, *77*, 393–404.
- Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. (2013). *The MOOC phenomenon: Who takes massive open online courses and why?* Working paper, University of Pennsylvania, Philadelphia.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*, 315–372.
- Elvers, G. C., Polzella, D. J., & Graetz, K. (2003). Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology*, *30*, 159–162.
- Evans, B., & Baker, R. (2016). MOOCs and persistence: Definitions and predictors. *New Directions in Institutional Research*, *167*, 69–85.
- Evans, B., Baker, R., & Dee, T. (2016). Persistence patterns in massive open online courses (MOOCs). *Journal of Higher Education*, *87*, 206–242.
- Frederick, S., Loewenstein, G., & O’Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*, 351–401.
- Gine, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, *2*, 213–225.
- Glass, C. R., Shikawa-Baklan, M. S., & Saltarelli, A. J. (2016). Who takes MOOCs? *New Directions in Institutional Research*, *167*, 41–55.
- Guàrdia, L., Maina, M., & Sangrà, A. (2013). MOOC design principles. A pedagogical approach from the learner’s perspective. *eLearning Papers*, *33*, 1–6.
- Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learning activity features. *eLearning Papers*, *37*, 7–16.
- Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*, 126–134.
- Haynie, D. (2015, May 27). Opportunity for credit a new hook for MOOCs. *U.S. News & World Report*. Retrieved from <http://www.usnews.com/education/online-education/articles/2015/05/27/chance-for-credit-gives-new-life-to-moocs>

- Hill, J. R. (2002). Overcoming obstacles in and creating connections: Community building in web-based learning environments. *Journal of Computing in Higher Education*, 14, 67–86.
- Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., 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 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2381263
- Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and reality. Full report*. New York, NY: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.
- Imbens, G. W., & Angrist, J. A. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475.
- Jordan, K. (2014). *MOOC completion rates: The data*. Retrieved from <http://www.katyjordan.com/MOOCproject.html>
- Kaur, S., Kremer, M., & Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123, 1227–1277.
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). *Recommending self-regulated learning strategies does not improve performance in a MOOC*. Working paper.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 443–478.
- Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, 107, 573–597.
- Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic Perspectives*, 3, 181–193.
- Macan, T. H., Shahani, C., Dipboye, R. L., & Phillips, A. P. (1990). College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology*, 82, 760–768.
- McCormick, A. (2003). Swirling and double-dipping: New patterns of student attendance and their implications for higher education. *New Directions for Higher Education*, 121, 13–24.
- Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, 56, 243–252.
- Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American Journal of Health Studies*, 16, 41–51.
- Nawrot, I., & Doucet, A. (2014). Building engagement for MOOC students: Introduction support for time management on online learning platforms. In *Proceedings of the 23rd International Conference on the World Wide Web* (pp. 1077–1082). New York, NY: ACM.
- Patterson, R. W. (2014). *Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course*. Working paper.
- Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs: Understanding the progression of users in massive open online courses. *Educational Researcher*, 43, 421–432.
- Qualtrics. (2013). *Qualtrics and instructor partner reveal top motivations for MOOC students*. Retrieved from <http://qualtrics.com/press/press-releases/qualtrics-and-instructure-partner-reveal-top-motivations-for-mooc-students/>
- Roper, A. R. (2007, November 1). How students develop online learning skills. *Educause Quarterly*, pp. 62–65.
- Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *Internet and Higher Education*, 7, 59–70.
- Thaler, R.H. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8, 201–207.
- Van Den Hurk, M. (2006). The relation between self-regulated strategies and individual study time, prepared participation and achievement in a problem-based curriculum. *Active Learning in Higher Education*, 7, 155–169.
- Veletsianos, G., & Shepherdson, P. (in press). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *International Review of Research in Open and Distributed Learning*.
- Ye, C., Kinnebrew, J. S., Biswas, G., Evans, B. J., Fisher, D. H., Narasimham, G., & Brady, K. A. (2015). Behavior prediction in MOOCs using higher granularity temporal information. In *Learning @ Scale 2015 Annual Conference Proceedings* (pp. 335–338). New York, NY: ACM.
- Zhang, D., Allon, G., & Van Mieghem, J. (2015). *Does social interaction improve learning? Field evidence from massive open online education*. Working paper, Northwestern University, Chicago, IL. Retrieved from http://www.kellogg.northwestern.edu/faculty/vanmieghem/htm/pubs/2015_MOOC_Zhang-Allon-VanMieghem_16Aug2015.pdf

Authors

RACHEL BAKER is an assistant professor of education policy at the University of California, Irvine, School of Education. She conducts research on education policy, student success, and student sorting in higher education.

BRENT EVANS is an assistant professor of public policy and higher education at Peabody College, Vanderbilt University. He conducts research on student success in higher education.

THOMAS DEE is a professor of education at Stanford University's Graduate School of Education. His research focuses largely on the use of quantitative methods (e.g., panel data techniques, instrumental variables, and random assignment) to inform contemporary policy debates.