Do Planning Prompts Increase Educational Success? Evidence from Randomized Controlled Trials in MOOCs*

Mark Andor[1], Katja Fels[1,2], Jan Renz [3] and Sylvi Rzepka[1]

[1]RWI - Leibniz Institute for Economic Research
[2]Ruhr-University Bochum
[3]Hasso Plattner Institute, University of Potsdam

January 22, 2019

Abstract

Massive Open Online Courses are a promising educational innovation. Yet, they suffer from high drop-out rates. As a remedy, we propose a planning prompt and test its effect on course completion and further outcomes such as course engagement and satisfaction in four large-scale randomized controlled trials. The results reveal an overall null effect on the completion rate, ruling out effect sizes beyond the [-7%, 3%] interval. However, this overall effect masks heterogeneity across and within courses: In one course the planning prompt increases course completion by 19%, highlighting the importance of replications in slightly different contexts. Using random causal forests, we also reveal tendencies for differential effects by subgroups. Better targeting could hence improve the effectiveness of planning prompts in online learning.

JEL-Classification: I21, I29, C93

Keywords: Massive Open Online Courses, planning prompt, behavioral economics, natural field experiment

*The authors would like to thank Franz Liedke and the entire openHPI-team for outstanding technical support throughout the project. Furthermore, we thank James Andreoni, Jeannette Brosig-Koch, John Haisken-DeNew, Andreas Gerster, David Jaeger, Anett John, Alexander Koch, Stefan Lamp, Jonathan Meer, Alfredo Paloyo, Michael K. Price, Nicolas Salamanca, Christoph M. Schmidt, Stefanie Schurer, Charles Sprenger, Marcus Tamm, Arne Uhlendorff, Matthias Westphal and participants at the the Annual Conference of VfS2018, ASSA 2018, Potsdam Empirical Economics Workshop 2017, EEA 2017, the M-BEES 2017, the AASLE 2017, and the BIEN annual conference 2016 for helpful comments and suggestions. The trial has been registered in the AEA RCT registry (RCT ID AEARCTR-0001780). Fels gratefully acknowledges the support of a special grant (Sondertatbestand) from the German Federal Ministry for Economic Affairs and Energy and the Ministry of Innovation, Science, and Research of the State of North Rhine-Westphalia. The e-mail addresses of authors are: mark.andor@rwi-essen.de, katja.fels@rwi-essen.de, jan.renz@hpi.de and sylvi.rzepka@rwi-essen.de.

1 Introduction

Following up intentions with actions is challenging. For education this is particularly true because it requires sustained effort, while the expected benefits may only occur in the distant future. Therefore, it is not surprising that Massive Open Online Courses (MOOCs), an innovative and promising educational tool for lifelong learning, suffer from high drop-out rates. Up to 75% of participants who intend to earn a certificate fail to do so (Reich, 2014). Despite low completion rates, MOOCs are disrupting lifelong learning and higher education. Up to now, they have reached 101 million learners globally (Shah, 2018). According to the Economist (Jan 12 2017) MOOCs make "balancing learning, working and family life" easier and quickly respond to new skill demands of the labor market with new courses or nanodegrees. An emerging literature documents and seeks to explain the massive disengagement in MOOCs. Student-related factors, such as experience, organizational skills and motivation, appear to be the most compelling causes for disengagement (Lee & Choi, 2011; Banerjee & Duflo, 2014; Kizilcec & Schneider, 2015). Yet, little is known about how to mitigate disengagement. So far, only few studies employ behaviorally-motivated interventions to address this challenge (e.g Martinez, 2014; Patterson, 2018; Kizilcec et al., 2016).

In this paper, we propose a planning prompt as a remedy for drop-out in MOOCs and test it in four large-scale randomized controlled trials (RCT). The RCTs were implemented in courses by openHPI and openSAP, two German MOOC-platforms operating internationally in the field of internet technology. The experiments classify as natural field experiments since participants were not informed about the study, the planning prompt was embedded in the learning platform as a pop-up, and course communication was carried on as usual (cf. Levitt & List, 2009).

Our paper contributes to the growing literature of behavioral economics in education (Levitt et al., 2016; Koch et al., 2015; Lavecchia et al., 2016; Damgaard & Skyt Nielsen, 2018). We test the nudge-like intervention of planning prompts, which previously has shown large positive effects in many other settings: it increased colonoscopy and mammography uptake (Milkman et al., 2013; Rutter et al., 2006), vaccinations (Milkman et al., 2011), savings (Lusardi et al., 2009), and voting (Nickerson & Rogers, 2010). The small body of experimental evidence on planning prompts in MOOCs has yielded mixed results. Baker et al. (2016) send their planning prompt via email with a link to a scheduling tool to a random sample of all students who enrolled at least two days prior to course start. They find small, significant negative effects. Yeomans & Reich (2017) place their planning prompt into a pre-course survey, ask open text questions about MOOC engagement plans, and find that it increases completion rates by 29%. Our study improves upon these two studies by treating all participants irrespective of enrollment timing or survey participation. This avoids self-selection into treatment. Furthermore, our planning prompt is embedded in the course interface as a pop-up rather than an external website or a survey question. Hence, it can be perceived as a regular feature of the platform. Consequently, our experimental design enables us to draw more general conclusions on how planning prompts work in online learning contexts.

Our paper further adds to the emerging literature on the transferability and scalability of experimental results to other contexts and other study populations (Allcott, 2015; Dehejia et al., 2015; Gechter, 2016; Al-Ubaydli et al., 2017b,a; Peters et al., 2018; Vivalt, 2015). By replicating the exact same treatment in four different courses with different study populations, our study follows the second level of replication as proposed by Levitt & List (2009).[1] This allows us to "dramatically increase" (Maniadis et al., 2014, p. 289) the ratio between true positive and false positive findings and avoid drawing false conclusions from a one-shot experimental setting.

Pooling all four courses, we find that the planning prompt has no significant effect on the overall certificate rate, course activity, and satisfaction with the course. This overall effect, however, masks substantial heterogeneity across and within courses. Analyzing the effects separately by course, we find that the planning prompt strongly increases completion rates in one course but produces null effects in the other three courses. The one positive effect may be a false positive. Yet, given the identical experimental design and overall course set up, it is also plausible that small details of the course design matter. For instance, the course with positive effects had relatively infrequent course communication. This detail could have enabled the planning prompt and its reminder to have such a positive influence on completion. Furthermore, the heterogeneity analysis within courses hints at opposite effects for different subgroups. The random causal forest algorithm suggests that these heterogeneous effects within courses are even more important than the course effects. It shows that for a skilled and committed group the planning prompt appears to be beneficial.

The overall null effects are in line with recent studies on planning prompts, which also find a precise zero (Carrera et al., 2018; Oreopoulos et al., 2018). These studies conclude that planning prompts may be effective to increase participation in one-shot actions but not in activities like exercising and studying, which require more sustained effort. Our heterogeneity analysis sheds light on other possible reasons for this overall null effect. Our results imply that planning prompts could be beneficial in courses that lack regular communication (as is

[1] According to Levitt & List (2009), there are three levels of replication: 1.) re-analyzing the original data from an experiment, 2.) conducting an independent new experiment under the same protocol but with different subjects, and 3.) applying a new research design suitable to test the validity of the first study's hypotheses.

the case in many self-paced courses) and for participants who engage early, are skilled, and share information about themselves. On the other hand, it seems to discourage participants who do not disclose information on education, skills, experience, or gender. This suggests that tailoring the planning prompt to the course context and targeting participants appears crucial to make planning prompts successful.

On a general stance, our results highlight the peril of drawing conclusions from just one RCT. Chronologically, the course with the strong positive effects was the first of our four RCTs. Had we relied on this single RCT, conclusions would have been very different and ultimately seriously misleading. Our results therefore empirically underline the calls for replications (Al-Ubaydli et al., 2017b,a).

The remainder of this paper is structured as follows. Section 2 describes the experimental set-up. Section 3 introduces the empirical strategy. Section 4 presents results from a pooled analysis and from a heterogeneity analysis across and within courses. Section 5 concludes and discusses the findings.

2 Experimental set-up

2.1 Context

We conduct four natural field experiments in MOOCs of openHPI and openSAP, two MOOC-providers offering courses in internet technology, computer science, and software usage and development. Concretely, the experiments took place in the courses "Linked Data Engineering" and "Web-Technologies", offered by openHPI, as well as "SAP Fiori for iOS – An Introduction" and "Getting Started with Data Science", offered by openSAP. Subsequently, we will refer to these courses as "Linked", "Web-tech", "Fiori", and "Data Science" respectively.[2]

The basic structure of all four courses is very similar. The two MOOC-providers have identical user interfaces except for the provider logo. Most courses are held in English. Only Web-tech was taught in German. Participating in a course and obtaining a certificate is free of charge with both providers; users register with their name and a valid email address. All courses consist of video-based instruction, ungraded quizzes, graded weekly assignments and a graded final exam. To earn a certificate, a so-called "Record of Achievement" (RoA), participants need to collect at least 50% of all possible course points via graded assignments

[2]Linked can be found online at: https://open.hpi.de/courses/semanticweb2016. The course consisted of six weeks of instruction between October 17, 2016 and December 12, 2016. Web-tech is available online at: https://open.hpi.de/courses/webtech2017. The course was taught for six lecture weeks between February 6, 2017 and April 7, 2017. The Fiori-course is available online at: https://open.sap.com/courses/ios1. It had three weeks of instruction between November 15, 2016 and December 14, 2016. Data Science can be accessed at: https://open.sap.com/courses/ds1. It was first taught between February 1, 2017, and March 23, 2017, and comprised six weeks of teaching.

and the final exam (Renz et al., 2016). Even though material can be studied at any time after its release, graded activities need to be completed by a certain deadline in order to earn points.[3]

2.2 Experimental design

We randomly assign participants into control and treatment groups with their first click on the course platform after the course started (Figure 1). The treatment is transmitted via a pop-up asking participants to determine the concrete time for their next MOOC-session. It appears very early in the course: after the third click on the platform, before a video starts (Figure 2). The pop-up also informs them that they will receive a reminder email shortly before their self-set study time. This reminder is in addition to the standard course emails which participants receive from the teaching team informing them about new material and activities in the course or the deadlines for the assignments or final exam. The treatment therefore is a combination of the planning prompt on the course platform and a reminder email.

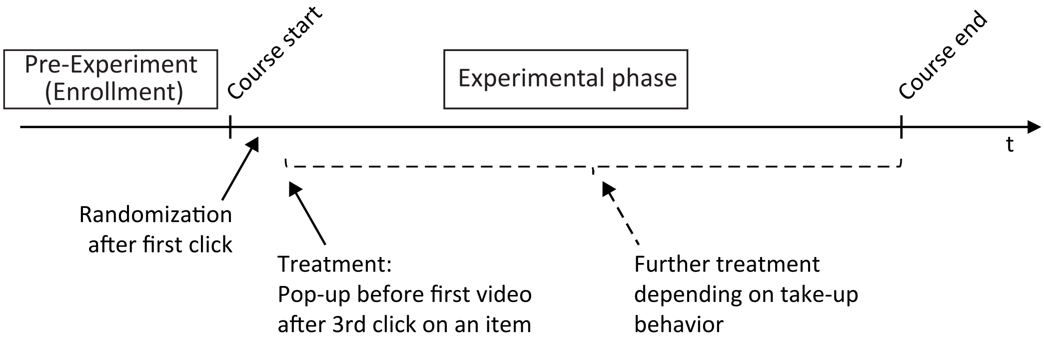


Figure 1: Timeline of the experiment

Using the planning prompt is optional since the treated participants can click “cancel” or close the pop-up without setting their next study time. These participants see the pop-up again in the following week. Participants, who decide to use the planning tool, set a time for their next session and click “schedule”. They receive the pop-up again in the following session.[4]

Because the treatment via pop-ups interrupts the learning flow, the control group also receives a pop-up in the very first session. It reads “Keep up the good work. Please press continue to watch the next video”. Figure 3 shows a screen shot of this pop-up.

[3]Please note that to meet the point requirement for the certificate, it is not necessary to hand in all assignments.
[4]A session is defined as continuous activity on the MOOC-platform without an interruption lasting longer than 30 minutes.

Figure 2: Treatment group pop-up

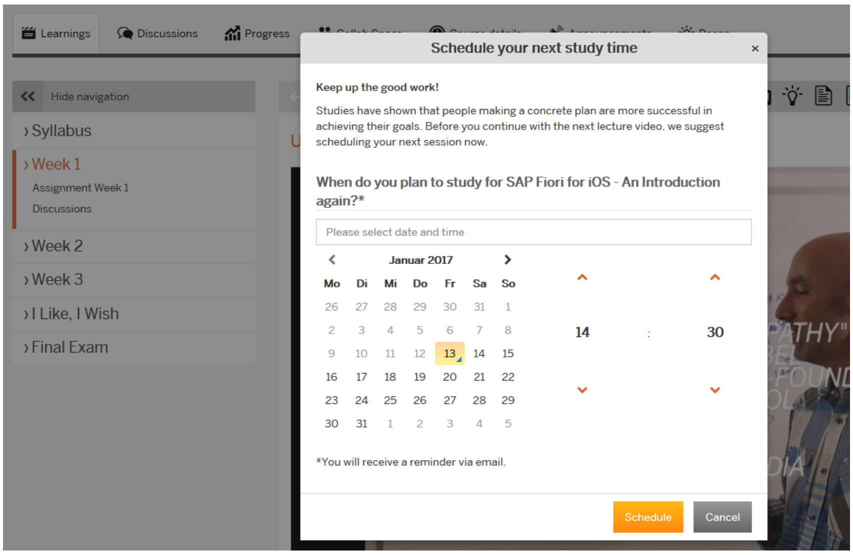


Figure 3: Control group pop-up

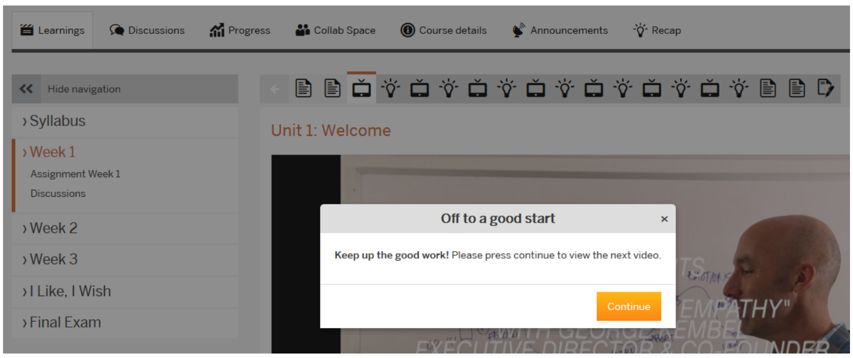


Figure 2 and Figure 3 illustrate that the treatment is embedded in the normal user interface and follows the corporate design of the MOOC-providers. This ensures that the planning prompt is perceived as a regular feature of the platform, clearly distinguishing our experimental set-up from previous studies. Furthermore, the study sample is not selective: all active MOOC-participants take part in the experiment. In contrast, previous studies have sent their planning prompt via email outside of course communication (Baker et al., 2016) or placed the planning prompt into a lengthy pre-course survey (Yeomans & Reich, 2017). In short, our experiment is a natural field experiment, where participants are unaware of the experimental nature of the intervention. This design allows us to draw a more general conclusion on the effect of planning prompts in MOOCs.

2.3 Potential channels

There are several channels through which the planning prompt can potentially affect behavior (Beshears et al., 2016). First, the planning prompt may provide more structure to the course helping participants spread their activity more evenly over the course of the MOOC. This consideration is based on the model by O'Donoghue & Rabin (2008). They suggest that individuals are more likely to procrastinate in long-term projects when they can flexibly choose when to work. This flexibility, they argue, paves the way for time-inconsistent individuals lacking self-control to put off work because hyperbolic discounting makes future work appear less cumbersome. In this sense, the planning prompt could have the effect of reducing flexibility and helping participants to overcome procrastination.

Second, scheduling the next study time can also be viewed as setting a goal and with it an internal reference point for future action (Koch & Nafziger, 2011). According to prospect theory, failing to study at the self-set time would create greater disutility than the utility realized when achieving the goal (Kahneman & Tversky, 2013). Viewed in this way, the planning prompt should therefore have a positive effect on course activity since individuals seek to maximize their utility.

Third, the planning prompt causes people to think about how to follow through with their intention making the required effort costs more salient at a very early stage of the course. For some individuals this may mean that they realize early on that they were over-optimistic about their availability. They may drop out earlier than they would have without the planning prompt (Beshears et al., 2016).

Finally, the reminder email of the self-set study time could affect MOOC completion positively by mitigating limited memory and inattention (Calzolari & Nardotto, 2016; Patterson, 2018). The reminder may help busy individuals to recall the original intention of studying the MOOC-material and hence increase the likelihood of earning a certificate.

In sum, this discussion of theoretical channels shows that there is reason to have a positive prior; nevertheless, theory also allows for an ambiguous overall outcome.

3 Empirical strategy

3.1 Data

Our sample consists of all enrolled participants, who click on at least three items, e.g. videos, quizzes, or reading material, of the course. The total pooled sample consists of 15,574 participants: 2,090 from the Linked course, 5,161 from Web-tech, 2,647 from Fiori, and 5,676 from Data Science.

For our analyses, we can use three types of data, which differ in how they are collected. Data is gathered either via the browser, the user's profile, or a survey. The browser collects information on the interaction with the platform, e.g. the time that the user is online, the number of videos played or quizzes submitted (Renz et al., 2016). It also picks up browser information such as the type of browser and the country from which the user accesses the MOOC-platform. This information collected by the browser is the most reliable information since it is elicited automatically and it is available for nearly every participant. Additional data on socio-economic characteristics is available for those participants who provide more information on their profile or in surveys. This type of information is likely to be selective and can only provide a non-representative overview of participants' characteristics.

Table 1 provides summary statistics for the pooled sample. As expected after randomization, the number of participants and all characteristics are well-balanced across the experimental groups. We only observe significant differences at a 5% level for one educational variable. Since the observed difference is very small in nature (0.6 percentage points) and participants with high-school education only represent a small minority of the sample (4%), this should not influence results.

Table 1, moreover, reveals that Germans make up the biggest participant group (39%) in the sample, followed by Indians (21%) and US-Americans (15%). Around two thirds of course participants (65%) enrolled before the course started. After the course started, the average participant logs in within 8 days (200 hours) for the first time. At least 9% of all participants are affiliated with SAP since they enrolled with a corporate e-mail address. About 22% of all participants use both the mobile app and a desktop device for course work. Judging from the non-missing socio-economic characteristics, the average participant appears to be a middle-aged man with university education: Only 7% of participants indicate being female, 23% are between 30-49 years old (those aged below 30 and above 50 make up another share of 9% each), and 19% report having a Master's degree or a PhD, while 12% have a

Table 1: Summary statistics

	(1) Control	(2) Treated	(3) Difference C & T
Panel A: Browser-based information			
Country			
Germany	0.3849	0.3832	0.0016
	(0.4866)	(0.4862)	(0.0078)
India	0.2072	0.2039	0.0032
	(0.4053)	(0.4029)	(0.0065)
US	0.1508	0.1518	-0.0010
	(0.3579)	(0.3589)	(0.0057)
Other country	0.4743	0.4834	-0.0091
	(0.4994)	(0.4998)	(0.0110)
Missing	0.0053	0.0064	-0.0011
	(0.0728)	(0.0800)	(0.0017)
Affiliated with SAP	0.0927	0.0859	0.0068
	(0.2900)	(0.2802)	(0.0046)
Enrolled prior to course start	0.6538	0.6526	0.0012
	(0.4758)	(0.4762)	(0.0076)
First login (in hours)	200.7243	201.8791	-1.1548
	(285.0794)	(285.5638)	(4.5730)
Mixed device user	0.2249	0.2147	0.0102
	(0.4176)	(0.4107)	(0.0066)
First-time user	0.2028	0.2032	-0.0004
	(0.4021)	(0.4024)	(0.0064)
Panel B: Profile-based information			
Female	0.0650	0.0615	0.0035
	(0.2466)	(0.2403)	(0.0039)
Gender missing	0.5792	0.5853	-0.0061
	(0.4937)	(0.4927)	(0.0079)
Age			
< 30	0.0857	0.0850	0.0007
	(0.2799)	(0.2789)	(0.0045)
$30-49$	0.2279	0.2216	0.0063
	(0.4195)	(0.4154)	(0.0067)
50+	0.0915	0.0949	-0.0034
	(0.2884)	(0.2931)	(0.0047)
Missing	0.5949	0.5985	-0.0036
	(0.4909)	(0.4902)	(0.0079)
Education			
High-school student	0.0382	0.0323	0.0059**
	(0.1916)	(0.1767)	(0.0030)
Bachelor	0.1202	0.1221	-0.0019
	(0.3252)	(0.3274)	(0.0052)
Master or PhD.	0.1937	0.1873	0.0064
	(0.3952)	(0.3902)	(0.0063)
Other or Missing	0.6480	0.6583	-0.0103
	(0.4776)	(0.4743)	(0.0076)
IT-skills			
Beginner	0.0706	0.0667	0.0039
	(0.2562)	(0.2495)	(0.0041)
Advanced	0.1830	0.1853	-0.0022
	(0.3867)	(0.3885)	(0.0062)
Expert	0.1242	0.1205	0.0038
	(0.3299)	(0.3255)	(0.0053)
Missing	0.6221	0.6276	-0.0054
	(0.4849)	(0.4835)	(0.0078)
Panel C: Survey-based information			
Answered pre-course survey	0.4541	0.4410	0.0130
	(0.4979)	(0.4966)	(0.0096)
Answered post-course survey	0.1837	0.1724	0.0112*
	(0.3872)	(0.3778)	(0.0061)
Course intention			
Earn COP	0.2465	0.2476	-0.0011
	(0.4311)	(0.4317)	(0.0124)
Earn ROA	0.6185	0.6139	0.0046
	(0.4859)	(0.4869)	(0.0140)
Browse	0.0974	0.0956	0.0018
	(0.2966)	(0.2941)	(0.0085)
Don't know yet	0.0367	0.0404	-0.0036
	(0.1882)	(0.1969)	(0.0055)
Professional interest	0.6309	0.6436	-0.0127
	(0.4827)	(0.4790)	(0.0138)
Impatient participant	0.2733	0.2583	0.0150
	(0.4458)	(0.4378)	(0.0127)
Observations	7704	7870	15574

Notes: Panel A shows summary statistics that are collected by the browser. Panel B variables primarily stem from profile entries. Panel C variables were elicited in pre- and post-course surveys. Panel B and C variables are only available for a non-representative sub-sample. "Affiliated with SAP" means participants enrolled with a corporate e-mail address or logged in through their SAP account. "First login" provides the time which elapsed between the course start and when participants begin being active in the course. For the classification of "impatient participant" see description in footnote 6. The variables "Course participation due to professional interest", "Course intention" and "Impatient participant" were collected for Web-tech and Data Science only. Columns 1 and 2 display the sample means with standard deviations in parentheses for the control and treatment group. Column 3 shows the differences between these two groups (control group mean − treatment group mean) and the corresponding standard error. The significance level of a t-test are indicated by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.
The question for course intention was asked slightly differently in both courses. The Data Science Course asked for these indicated categories directly. For the Web-tech course we interpreted "engaging with the course material by reading the material and doing the tests to earn a certificate at the end" as "Earn ROA", "learning the course topics by watching videos" as "Earn COP", and "None of the above" as "Browse".

Bachelor's degree. However, gender and education information is missing for 60 to 65% of all participants, therefore, these numbers may not be very representative of the overall sample.

The survey-based information reveal that 63% of the sample have a professional interest in the course; 62% aim to earn a certificate (ROA) and another 25% want to complete the course with a confirmation of participation (COP)[5]. A minority (10%) of the sample participates to browse or does not know yet which course outcome they intend (4%). Finally, more than one fourth of participants (27%) assesses themselves as impatient.[6]

3.2 Usage of the planning tool

Around 30% of all treated participants used the planning tool to schedule their next study time at least once. We define intentional usage of the planning tool conservatively as scheduling the next study time at least once with the scheduled time being at least 2 hours in the future. OUr take-up rate is a much higher than the 13% in Baker et al. (2016), a previous planning prompt intervention in a MOOC context. It is also at the upper bound of take-up rates of comparable interventions such as commitment contracts for savings or exercise which range between 11% and 28% (Royer et al., 2015).

Figure 4a shows how MOOC-participants schedule their next study time. The largest spikes are just one day ahead. About 49% of participants who use the tool schedule a time between 21 and 27 hours into the future. There is also bunching at other multiples of 24 indicating that people prefer to schedule the same time of the day a few days ahead. Many participants also set their next study time sooner than just the next day. 19% of all participants who use the tool schedule a time 3 to 19 hours into the future.

Figure 4b depicts the difference between the scheduled and the actual time MOOC-participants return to the platform. Here the spike around zero indicates that many MOOC-participants return for their next study session at about the time they scheduled. For 30% of MOOC-participants their next session lies within a 3-hour time window around their scheduled time. 24% of MOOC-participants who set a time logged in again 25 to 5 hours ahead of their scheduled time, and 19% log in 4 to 25 hours after their scheduled time. Furthermore, we can again see some bunching at multiples of 24h suggesting that the participants return at the time they intended though on a following day.

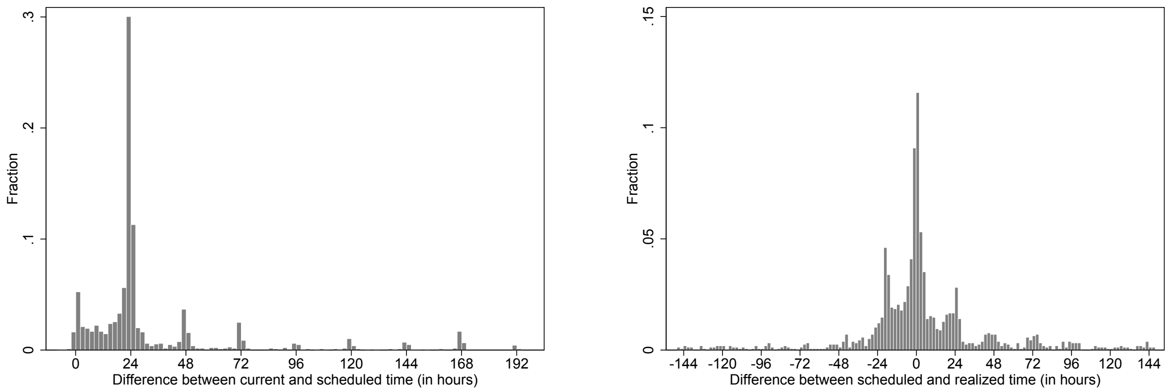
[5] For a ROA, participants need to earn more than 50% of all possible course points. A COP is issued to those participants who have viewed at least 50% of the course material.

[6] We asked participants to answer "Are you generally an impatient person or someone who always shows great patience?" on a Likert-scale from 0 [very impatient] to 10 [very patient]. Participants who chose the categories 0-4 are classified as impatient. The question has been used before in the SOEP Vischer et al. (2013).

Figure 4: Planning tool usage

(a) Difference between current time at scheduling and scheduled time

(b) Difference between scheduled time and realized time



Notes: The Figure displays time differences in hours. Bin width is two hours. Observations with differences greater than 200 (a) or 150 (b) hours are cropped to ease visibility.

3.3 Estimation strategy

To estimate the effect of the planning prompt on course completion and course activity, we employ the following regression

$$Y_i = \alpha + \beta T_i + \varepsilon_i \,,$$

where Y_i stands for course completion, intermediary outcomes such as video plays, number of sessions, session duration, number of quizzes submitted, and total points, or other outcomes such as course satisfaction and stress levels of individual i. T_i indicates the treatment status of individual i. It takes on the value 1 when the participant is selected into the group that sees the planning prompt. The β coefficient provides the causal effect of the planning prompt irrespective of whether they actually interact with the planning tool or not. It is the coefficient of interest.[7] ε_i captures the remaining idiosyncratic error.

4 Results

4.1 Pooled analysis

The regression results presented in Table 2 (column 1) for the pooled sample of all four courses reveal that the treatment had no significant effect on the certification rate. The 95%-

[7]We view the planning prompt as the treatment. However, one could also consider planning tool usage to be the treatment. Then, to estimate the treatment effect on the treated one would instrument planning tool usage with the treatment assignment. Results of such two-stage least squares estimations are presented in Table A1. The point estimates in the second stage are not estimated with more precision; this is why we only report our original treatment definition in the text.

confidence-interval suggests that potential effect sizes lie between -7% and 3%. Compared to previous studies, these are very small effect sizes in both directions. Baker et al. (2016) detect negative treatment effects of 52% compared to the control group completion rate and Yeomans & Reich (2017) report a positive effect size of 29%. In this respect, our results represent a tightly estimated null overall effect.

Similarly, the planning prompt does not significantly impact intermediate outcomes which measure course engagement (Table 2, Columns 2-6). The point estimates and the effect sizes of the confidence intervals suggest tendencies towards a more frequent and longer study duration and more videos played by treated participants. Yet, treated participants have a tendency to earn fewer points and submit less self-test quizzes. Hence, the point estimates suggest they lag behind on more demanding course activities.

Finally, we do not find that the planning prompt had negative side effects (see Table 2 columns 7 and 8). Previous literature shows that using nudge-like interventions may also have disadvantageous effects on overall welfare even when desired positive effects on the main outcome variable can be detected (c.f. Damgaard & Gravert, 2018; Allcott & Kessler, 2018). In our case, such disadvantageous effects could be putting more stress on participants or decreasing course satisfaction. Therefore, we asked participants about perceived stress levels and course satisfaction in a post-course survey.[8] Overall course satisfaction appears to be unaffected by the planning prompt. The point estimate for perceived stress levels indicates that course participants in the treatment group tend to be less stressed than the control group. However, the coefficient is not statistically significant. We therefore conclude that the planning prompt did not bring about a worse course experience for the treated who complete the course.

[8]Note, that these results are based on a selective sample of those participants who remained active in the course until the last week and answered the post-course survey: these were 282 participants (13.49%) in Linked, 984 (18.86%) in Web-tech, 612 (23.12%) in Fiori, and 893 (17.20%) in Data Science. As this caveat applies to the treatment group as well as to the control group the comparison of the two groups seems viable.

Table 2: Effects on certification rate and intermediate outcomes

	(1) Certificate	(2) Sessions	(3) Duration	(4) Points	(5) Quizzes	(6) Videos	(7) Satisfaction	(8) Stress
Treatment	-0.005	0.066	0.002	-2.075	-0.558	0.078	0.001	-0.032
	(0.007)	(0.418)	(0.187)	(1.428)	(0.449)	(2.057)	(0.017)	(0.020)
Constant	0.299***	14.507***	5.290***	84.422***	19.135***	47.135***	0.503***	0.522***
	(0.007)	(0.320)	(0.135)	(1.783)	(0.377)	(1.994)	(0.019)	(0.020)
CI effect sizes	[-7%, 3%]	[-4%, 6%]	[-4%,6%]	[-8%,1%]	[-6%,2%]	[-8%,9%]	[-3%,3%]	[-7%,7%]
N	15574	15574	15574	15574	15574	15574	2679	2553

Note: Results are obtained from an OLS regression. The outcome in column 1 is the share of certificate earners, in column (2) the number of sessions (defined as period of interaction with the MOOC platform that is not interrupted for longer than 30 minutes), in column 3 the total duration measured in hours, in column 4 total points earned with the assignments and/or the final exam, column 5 the number of ungraded quizzes submitted, and in column 6 the number of videos played. In column 7 and 8, satisfaction and perceived stress levels are depicted. They were elicited in a post-course survey on an 11-point Likert scale, 0-4 are aggregated to built the indicator variables used as regression outcomes. CI effect sizes refer to the 95%-confidence interval (CI). Course fixed effects included. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Taking the effect sizes at face value, the pooled findings give rise to the interpretation that if the planning prompt affects MOOC completion or engagement at all, the overall direction is likely to be negative. By this our more general experimental design confirms the results of previous planning prompt interventions in MOOCs which focused on positively selected samples and did not embed the treatment in the course interface. If anything, Baker et al. (2016) find that planning prompts tend to reduce MOOC engagement and the probability of earning a certificate. While Yeomans & Reich (2017)'s main result highlights that asking students to verbalize their study plans significantly increases MOOC completion rates, they also reveal that plans which focus on time aspects are not successful. This provides suggestive evidence that operationalizing planning prompts by asking for specific study times, like our study and that of Baker et al. (2016), may not be a promising strategy for MOOCs. Oreopoulos et al. (2018) draw a similar conclusion after finding a zero effect on nudging college students to schedule their time for studying.

4.2 Heterogeneity analysis

4.2.1 Heterogeneity across courses

The small and insignificant point estimates might be the result of offsetting effects. In other words: the planning prompt may affect certain subgroups of participants in opposing directions, resulting in an overall zero. Studies in domains other than education have highlighted that effects for subgroups can substantially differ from another (John, 2015; Byrne et al., 2018). Therefore in the following, we explore heterogeneous treatment effects across courses.

Estimating the treatment effects separately for each course reveals substantial heterogeneity across courses (Table 3 Column 1). Most notably, in the Linked course the planning prompt affected the certification rate significantly positively. In Linked, the certification rate was raised by 3.4 percentage points, a relative increase of 19% compared to the control group mean.[9] For the other three courses, coefficients are small, negative, and statistically insignificant. The confidence intervals imply that any meaningful effect sizes are more likely to be negative.

These strong heterogeneities are also visible in MOOC engagement. Treated participants in the Linked course show higher course engagement for all intermediary outcomes (Table 3 Columns 2-6). In terms of effect size and significance the number of videos played appears to be very important: Treated participants watched about 11 (24.5%) videos more than participants from the control group. For the other three courses, there is further tentative evidence for adverse effects of the planning prompt on MOOC engagement. The point estimates and

[9] Baseline certification rates in the four courses differed, ranging from 18% in Linked to 25% in Web-tech, 30% in Data Science, and 36% in Fiori.

effect sizes implied by the 95%-confidence intervals suggest that treated participants tend to perform worse than the control group in nearly all engagement indicators (Table 3).

While our prior was that the planning prompt would nudge participants positively, the confidence intervals suggest that negative effects are more likely. For a number of other domains, a nascent literature points out that nudge-like interventions such as planning prompts can have disadvantageous effects: among them fundraising (Damgaard & Gravert, 2018), savings (John, 2015), taxation (Dwenger et al., 2016), and energy conservation (Schultz et al., 2007). Our study adds another domain to this literature: behaviorally motivated interventions can also potentially dissuade participants from engaging in online courses.

Table 3: Effect of planning prompt on intermediary outcomes (by course)

	(1) Certificate	(2) Sessions	(3) Duration	(4) Points	(5) Quizzes	(6) Videos
Panel A: Linked						
Treated	0.034*	1.143	0.752	2.867	1.890	11.497*
	(0.017)	(1.197)	(0.547)	(2.369)	(1.279)	(5.897)
Constant	0.178***	18.210***	7.473***	30.987***	20.499***	46.519***
	(0.012)	(0.813)	(0.350)	(1.622)	(0.877)	(2.811)
CI effect sizes	[0;38%]	[-7%;19%]	[-4%;24%]	[-6%;24%]	[-3%;21%]	[0%;50%]
N	2090	2090	2090	2090	2090	2090
Panel B: Fiori						
Treated	-0.018	-0.330	-0.240*	-2.899	-0.897	-4.335**
	(0.019)	(0.340)	(0.135)	(2.792)	(0.555)	(1.851)
Constant	0.355***	8.882***	3.173***	62.909***	15.153***	25.728***
	(0.013)	(0.243)	(0.106)	(1.980)	(0.399)	(1.478)
CI effect sizes	[-16%; 5%]	[-11%;4%]	[-16%;1%]	[-13%;4%]	[-13%;1%]	[-31%;-3%]
N	2647	2647	2647	2647	2647	2647
Panel C: Web-tech						
Treated	-0.013	0.088	0.060	-0.816	-1.139	-4.367
	(0.012)	(1.022)	(0.467)	(1.734)	(1.025)	(3.759)
Constant	0.303***	26.954***	12.001***	43.178***	33.222***	69.518***
	(0.009)	(0.726)	(0.330)	(1.247)	(0.749)	(2.796)
CI effect sizes	[-12%; 3%]	[-7%;8%]	[-7%;8%]	[-10%;6%]	[-9%;3%]	[-17%;4%]
N	5161	5161	5161	5161	5161	5161
Panel D: Data Science						
Treated	-0.006	-0.167	-0.215	-4.655	-0.773	1.972
	(0.012)	(0.484)	(0.194)	(3.227)	(0.603)	(3.835)
Constant	0.250***	14.624***	5.399***	85.724***	19.244***	46.180***
	(0.009)	(0.341)	(0.137)	(2.325)	(0.428)	(2.291)
CI effect sizes	[-12%; 7%]	[-8%;5%]	[-11%;3%]	[-13%;2%]	[-10%;2%]	[-12%;21%]
N	5676	5676	5676	5676	5676	5676

Notes: Panel A provides OLS estimates for the Linked course; Panel B for the Fiori course; Panel C for the Web-technologies course; Panel D for the Data Science course. The outcome in column 1 is the share of certificate earners, in column (2) the number of sessions defined as period of interaction with the MOOC platform that is not interrupted for longer than 30 minutes, in column 3 the total duration measured in hours, in column 4 total points earned with the assignments or the final exam, column 5 the number of ungraded quizzes submitted, and in column 6 the number of videos played. Robust standard errors in parentheses. CI effect sizes refer to the 95%-confidence interval (CI). *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

What drives the heterogeneous effects across courses? While the observed positive effects in the Linked course could be a false positive, it is also plausible that these effects are associated with details of the course structure. All openHPI and openSAP courses have the same format. Therefore, the Linked course does not differ much in most observable course characteristics, such as duration, video time, number of quizzes (see characteristics and Z-

Scores in Table 4). Yet, it deviates from the mean of all courses in the number of email alerts per week. Due to a platform error only half of the usually standard amount of the emails from the teaching team were sent out successfully. This makes the Linked course resemble a self-paced course, which often has less frequent communication. However, treated participants who scheduled their next study time did receive a reminder email two hours before their self-set study time including a link to new material (Figure A1). Hence, for treated participants in the Linked course, the planning prompt and its reminder substantially raised the frequency of emails relative to the control group. We conclude that such details of course structure matter: while the planning prompt and its reminder can yield substantial positive effects in a context without a frequent email stream, they appear less beneficial in contexts with frequent emails. This may be because the effect on the marginal participant is likely to be higher in a context of infrequent communication (Coffman et al., 2015).

Table 4: Course characteristics

Course characteristics	Linked	Web-tech	Fiori	Data Science
Duration in weeks	6 (0.7)	6 (0.7)	3 (-2.0)	6 (0.7)
Total video duration in hours per week	2.0 (0.6)	2.4 (1.4)	1.1 (-0.9)	1.0 (-1.1)
Total number of quizzes per week	7.2 (0.9)	7.3 (1.1)	6.3 (-0.2)	5.2 (-1.8)
Total number of quiz questions per week	16.2 (0.2)	22.5 (1.8)	13.3 (-0.6)	10.0 (-1.4)
Total number of assignments per week	1 (0)	1 (0)	1 (0)	1 (0)
emails per week (control group)	0.7 (-2.0)	1.5 (0)	2.3 (2.0)	1.5 (0.0)
Number of surveys	1 (-1)	2 (1)	1 (-1)	2 (1)

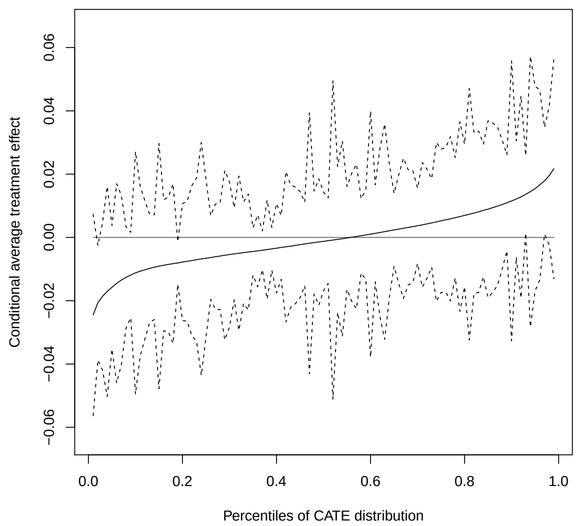
Notes: The Table shows means of the course characteristics and the corresponding z-scores in parentheses which standardize each courses characteristics by the mean of all four courses.

4.2.2 Heterogeneity within courses

The overall insignificant effect of the planning prompt on course completion may also be brought about by opposing effects on different subgroups of participants within courses. To uncover heterogeneous treatment effects we rely on the causal random forest method proposed by Wager & Athey (2018). It provides a data-driven way to identify subgroups with different treatment effects. They introduce the concept of "honesty" as a remedy for overfitting. "Honesty" is achieved by using independent samples for covariate selection and for the actual estimation of the heterogeneous effects. First, a random causal forest is trained with one random subsample. Second, another random subsample is used to estimate the heterogeneous effects. These estimates are unbiased and provide valid confidence intervals.

Training the causal forest, we use the optimal parameters determined by the built-in tuning function in the "grf"-package in R. We grow forests with the advised number of 4000 trees. We include all covariates that are pre-determined or elicited early in the course as

Figure 5: Conditional average treatment effect



Notes: This figure displays the distribution of the conditional average treatment effect in percentage points (y-axis) for the all percentiles of MOOC participants (x-axis). The dashed lines indicate the 90% confidence intervals. It is estimated using the causal forest function of the "grf" R-package. We use the built-in tuning function to specify the parameters and grow 4000 trees.

potential candidates for heterogeneous effects.[10] We also include indicator variables for the four different courses. This will allow us to determine whether heterogeneities are more pertinent across courses or across participant subgroups.

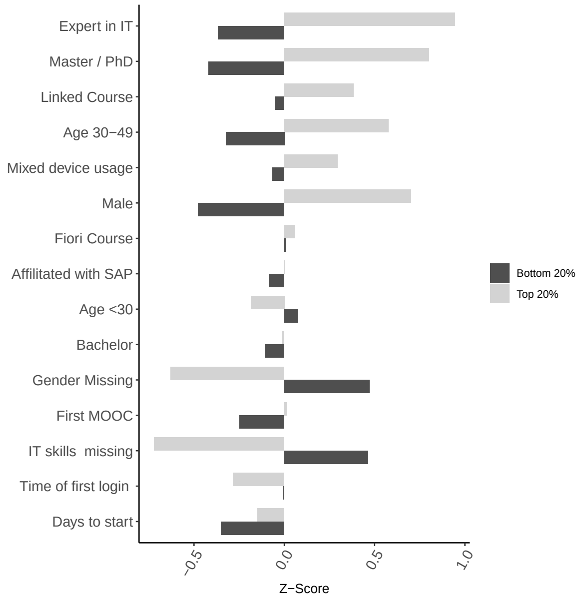
Figure 5 displays the distribution of the conditional average treatment effect (CATE). The point estimates suggest that the effect varies between −3.5 and 3.1 percentage points. While the confidence intervals are wide and only reach significant levels at the outer edges of the distribution, Figure 5 still provides clear evidence for effect heterogeneity.

Who are the participants at the two ends of the distribution? Figure 6 characterizes participants in the bottom 20% and the top 20% of the CATE distribution. We plot the z-scores of the 15 most important characteristics which were determined while training the random forest (see Table A2).[11] To provide some intuition, we investigate the share of single features at the two ends of the distribution. In practice, it will be combinations of features that are relevant. As Li & Baker (2018) point out, there are different types of MOOC-participants which have distinct engagement and outcome patterns. For example,

[10]All variables listed in Table 1 are included except for the indicator of post-course survey participation because this variable is elicited very late in the course. Two variables enter into the random forest estimation in a different form: For countries we use indicator variables for all countries that have at least 40 observations. 38 countries comply with this criteria. Enrollment timing is included as a continuous variable measured in days.

[11]These are the variables at which the trees of the random forest were split often. This identifies the characteristics which increase the prediction accuracy.

Figure 6: Characteristics of the bottom and top of the CATE distribution



Notes: This figure displays the Z-scores ($\frac{(\bar{x}_s - \bar{x}_p)}{\sigma_p}$ where s indicates the subsample and p the entire sample) of the characteristics of the participant groups for the bottom 20% and the top 20% of the CATE distribution. We only show the top 15 covariates which the random forest algorithm deemed most important because they lower the prediction error. See the notes to Table 1 for a description of the variables.

the first course indicator has rank 5. This suggests that individual characteristics are more important, but the interaction between course effects and individual characteristics is also not negligible.

It becomes evident which participants tend to benefit from the planning prompt the most: IT-experts, post-graduates, men, participants of working age, those who log in soon after the course started and use several devices to work on course material. All these groups are over-represented in the top quintile the CATE distribution. Our analysis also confirms the positive effect of the planning prompt for participants of the Linked course. Yet, it points out that other individual participant characteristics are more dominant in the top quintile. For instance, the share of IT-experts in the bottom of the distribution is 0.4 standard deviations lower than the population mean and in the top of the distribution it is 0.9 standard deviations higher than the population mean. The indicator "expert in IT" not only suggests that people of a high skill level react positively to the planning prompt; it may also show a higher commitment of those who choose to share this information on their profile or in a survey. The latter also applies to all other survey- or profile-based information (education, gender, age) which are optional to share. Further, at the top of the distribution, participants have a shorter time-lag between their first log-in and course start. These characteristics suggest that the planning prompt is most effective for participants who show skills, motivation, and commitment early on. Hence, those who benefit from the planning prompt, appear to be a positively selected group.

It is less clear which participants are dissuaded by the planning prompt. Only two characteristics stand out: Those who do not indicate their IT skills or their gender are over-represented by about 0.5 standard deviations at the bottom part of the distribution. They do not seem to benefit from the planning tool. Following the argumentation from above, we interpret this non-disclosure of information as a signal of less commitment to the course.

For some characteristics Figure 6 shows only small and sometimes non-linear heterogeneous effects. This suggests that these characteristics are important to determine the treatment effect in combination with other variables. Therefore, a ceteris paribus interpretation may not tell the whole story. For example, participants who state that they have a Bachelor's degree or who enroll using an SAP-affiliation are nearly evenly spread across the CATE distribution. All things equal, this means that the planning prompt does not help these participants more than others, however, in combination with other characteristics this may be the case. First-time MOOC participants are less present in the bottom part of the distribution, they appear more in middle. For them, the planning prompt's effect is close to zero. Finally, there are no clear heterogeneous effects with respect to enrollment timing: both at the bottom and at the top of the distribution participants who enroll early are over-represented. While Banerjee & Duflo (2014) suggest that participants who enroll before the MOOC starts are better in self-organization and perform better in the course, our results imply that the planning prompt may encourage but also discourage this group.

In sum, our analysis reveals there are heterogeneities. We conclude that the overall zero effect can be due to off-setting effects for different subgroups. The planning prompt appears most effective for a positively selected group, most notably participants who share information about themselves, show motivation by using several devices to complete the course work or a combination of these characteristics. Targeting this group, which is over-represented in the top quintile of the distribution, could yield treatment effects of up to 1.3 percentage points, which translates into an increase of about 5% compared to the control group.

5 Conclusion

This paper examines whether prompting participants to schedule their next study time in a MOOC increases certification rates. Based on four large-scale randomized control trials, we show that the planning prompt has a null overall effect on course completion and course engagement. It also has no significant negative side effects on satisfaction or stress levels. Yet, there are substantial heterogeneities across and within courses. In one course, treated participants have a 19% higher probability of completing the course with a certificate. The other three courses show no significant effects. The positive effect could be a false positive. However, we provide suggestive evidence that specific details of the course structure, such as the frequency of out-of-course communication, make a difference because they determine how much the marginal participant can be influenced. Furthermore, using the causal random forest algorithm (Wager & Athey, 2018), we reveal off-setting heterogeneous effects based on individual characteristics which matter even more than differences across courses. For a positively selected group of participants, the prompt has beneficial effects. These participants publicly state that they have a high skill level, are in a working-age, or male. In contrast, participants who do not disclose information about themselves, are negatively affected by the planning prompt. This may be explained by less commitment to the course in the first place.

Our results have important implications for two strands of the literature. First, our study highlights that interventions such as planning prompts motivated by behavioral economics provide no silver bullet in online education. Our experimental design allows for more general conclusions of planning prompts in MOOCs than previous studies since we implement the planning prompt directly in the course rather than outside of the course platform. We furthermore target all participants instead of just those who participate in a survey (Baker et al., 2016; Yeomans & Reich, 2017). Still, we detect no significant overall effect of the planning prompt on MOOC-participants. This casts further doubt on whether planning prompts are powerful enough to assist in domains where sustained effort is needed (Carrera et al., 2018; Oreopoulos et al., 2018). Furthermore, in education, online and offline, there is uncertainty about the best learning strategy. Hence, focusing on just the time input may not be the appropriate way to increase outcomes (Oreopoulos et al., 2018). Finally, the heterogeneity analysis implies that a planning prompt should only target subgroups who are likely to benefit. A promising avenue for further research is to investigate whether the subgroups identified by our study carry over to other nudges and to other domains.

Second, on a more general note, we complement the literature by showing that even in very similar contexts the transferability of causal effects across settings is limited (Al-Ubaydli et al., 2017b). In our case, it is plausible that a detail like the frequency of email communication can influence signs and effect sizes substantially. Had we extrapolated the findings from our first study, conclusions would have been misleading. We therefore recommend integrating replications even in the first study set-up, especially when the costs of doing so are low, as is the case in online education.

References

Al-Ubaydli, O., List, J. A., LoRe, D., & Suskind, D. (2017a). Scaling for economists: Lessons from the non-adherence problem in the medical literature. *Journal of Economic Perspectives*, *31*(4), 125–44.

Al-Ubaydli, O., List, J. A., & Suskind, D. L. (2017b). What can we learn from experiments? Understanding the Threats to the Scalability of Experimental Results. *American Economic Review*, *107*(5), 282–86.

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, *130*(3), 1117–1165.

Allcott, H., & Kessler, J. B. (2018). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics*, *forthcoming*.

Baker, R., Evans, B., & Dee, T. (2016). A randomized experiment testing the efficacy of a scheduling nudge in a massive open online course. *AERA Open*, *2*(4), https://doi.org/10.1177/2332858416674007.

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

Beshears, J., Milkman, K. L., & Schwartzstein, J. (2016). Beyond beta-delta: The emerging economics of personal plans. *American Economic Review*, *106*(5), 430–34.

Byrne, D. P., Nauze, A. L., & Martin, L. A. (2018). Tell me something i don't already know: Informedness and the impact of information programs. *The Review of Economics and Statistics*, *100*(3), 510–527.

Calzolari, G., & Nardotto, M. (2016). Effective reminders. *Management Science*, *63*(9), 2915–2932.

Carrera, M., Royer, H., Stehr, M. F., Sydnor, J. R., & Taubinsky, D. (2018). The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise. *NBER Working Paper Series*, (w24959).

Coffman, L., Featherstone, C. R., & Kessler, J. B. (2015). A model of information nudges. URL https://pdfs.semanticscholar.org/53ad/dcb393a4a770e4f307eadbea5716a74f7b9f.pdf

Damgaard, M. T., & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, *157*, 15–26.

Damgaard, M. T., & Skyt Nielsen, H. (2018). Nudging in education. *Economics of Education Review*, *64*, 313 – 342.

Dehejia, R., Pop-Eleches, C., & Samii, C. (2015). From Local to Global: External Validity in a Fertility Natural Experiment. *NBER Working Paper Series*, (w21459).

Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, *8*(3), 203–32.

Economist (Jan 12 2017). Learning and earning: The return of the MOOC: Established education providers v new contenders. *Economist*, *2017*.
URL http://www.economist.com/news/special-report/21714173-alternative-providers-education-must-solve-problems-cost-and

Gechter, M. (2016). Generalizing the results from social experiments: Theory and evidence from mexico and india.
URL http://www.personal.psu.edu/mdg5396/Gechter_Generalizing_Social_Experiments.pdf(accessedJune6,2017)

John, A. (2015). When commitment fails - Evidence from a regular saver product in the Philippines. *EOPP Discussion Papers*, *55*.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, (pp. 99–127). World Scientific.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Recommending Self-Regulated Learning Strategies Does Not Improve Performance in a MOOC. In J. Haywood, V. Aleven, J. Kay, & I. Roll (Eds.) *The Third ACM Conference*, (pp. 101–104).

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a Lens to Understand Online Learners. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1–24.

Koch, A., Nafziger, J., & Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic Behavior & Organization*, (115), 3–17.

Koch, A. K., & Nafziger, J. (2011). Self-regulation through Goal Setting. *The Scandinavian Journal of Economics*, *113*(1), 212–227.

Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, vol. 5, (pp. 1–74). Elsevier.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, *59*(5), 593–618.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, *8*(4), 183–219.

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in massive open online courses. *Computers Education*, *127*, 41 – 65.

Lusardi, A., Keller, P. A., & Keller, A. M. (2009). New ways to make people save: A social marketing approach. Tech. rep., National Bureau of Economic Research.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, *104*(1), 277–90.

Martinez, I. (2014). Never put it off till tomorrow? In *MOOCs as a Massive Research Laboratory*, (pp. 75–97).
URL http://libra.virginia.edu/catalog/libra-oa:7574

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415–10420.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2013). Planning prompts as a means of increasing preventive screening rates. *Prev Med*, *56*(1), 92–3.

Nickerson, D. W., & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science*, *21*(2), 194–199.

O'Donoghue, T., & Rabin, M. (2008). Procrastination on long-term projects. *Journal of Economic Behavior & Organization*, *66*(2), 161–175.

Oreopoulos, P., Patterson, R. W., Petronijevic, U., & Pope, N. G. (2018). Lack of study time is the problem, but what is the solution? unsuccessful attempts to help traditional and online college students. *NBER Working Paper Series*, (w25036).

Patterson, R. W. (2018). Can behavioral tools improve online student outcomes? experimental evidence from a massive open online course. *Journal of Economic Behavior & Organization*, *153*, 293 – 321.

Peters, J., Langbein, J., & Roberts, G. (2018). Generalization in the tropics - development policy, randomized controlled trials, and external validity. *World Bank Research Observer*, *33*(1), 34 – 64.

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

Renz, J., Schwerer, F., & Meinel, C. (2016). openSAP: Evaluating xMOOC usage and challenges for scalable and open enterprise education. In *The International Conference on E-Learning in the Workplace 2016*, (pp. 1–6).
URL http://www.icelw.org/proceedings/2016/ICELW2016/Papers/Renz_Schwerer_Meinel.pdf

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, *7*(3), 51–84.

Rutter, D. R., Steadman, L., & Quine, L. (2006). An implementation intentions intervention to increase uptake of mammography. *Annals of Behavioral Medicine*, *32*(2), 127–134.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*(5), 429–434.

Shah, D. (2018). By the numbers: Moocs in 2018. *Class Central*, *December 11*.
URL https://www.class-central.com/report/mooc-stats-2018/(accessed:January10,2019)

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, *120*(2), 142–145.

Vivalt, E. (2015). Heterogeneous treatment effects in impact evaluation. *American Economic Review*, *105*(5), 467–70.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Yeomans, M., & Reich, J. (2017). Planning prompts increase and forecast course completion in massive open online courses. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, (pp. 464–473). ACM.

Appendix

Figure A1: Reminder email of self-set study time

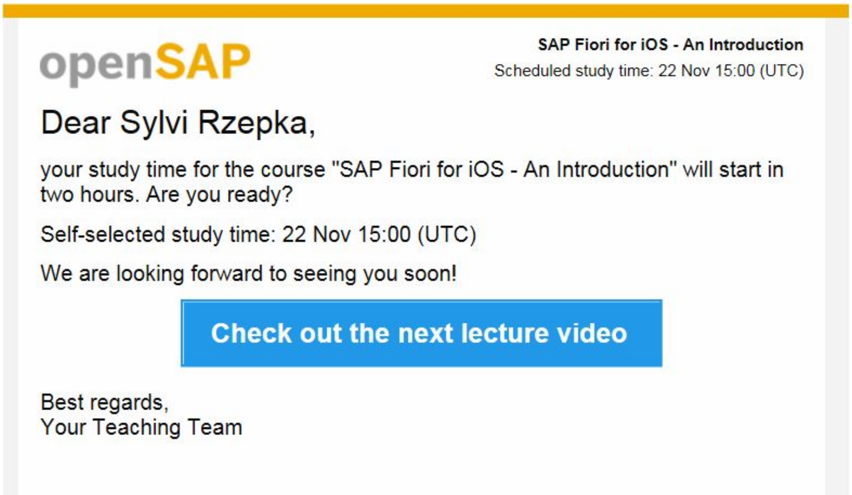


Table A1: Two-stage least square regressions on certificate completion

	(1) 1st-stage	(2) 2nd stage
Treatment assignment	0.298***	
	(0.005)	
Planning tool used at least once		-0.017
		(0.024)
Constant	-0.000***	0.277***
	(0.000)	(0.005)
CI effect sizes		[-26%, 10%]
Observations	15574	15574

Note: Results are obtained from a two-stage least square regression. CI effect sizes refer to the 95%-confidence interval (CI). Course fixed effects included. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A2: Top 15 characteristics

Rank	Characteristic
1	Time of first login
2	Days to start
3	Expert in IT
4	Master / PhD
5	Linked Course
6	Male
7	Age 30-49
8	Mixed device usage
9	IT skills missing
10	Bachelor
11	Age <30
12	Fiori Course
13	Gender Missing
14	Affilitated with SAP
15	First MOOC

Note: Ranking based on importance calculated by the causal random forest algorithm using the "grf"-package in R.