PRELIMINARY PLEASE DO NOT CITE OR CIRCULATE

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

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

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

September 30, 2018

Abstract

Massive Open Online Courses are a promising educational innovation but 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 suggest no significant positive effect on the overall completion rate. However, the overall effect masks heterogeneity across and within courses. In one course we find the planning prompt increases course completion by 19%. This highlights the importance of replications in slightly different contexts. Furthermore, we reveal tendencies for differential effects by subgroups.

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, Katja Görlitz, John Haisken-DeNew, David Jaeger, Alexander Koch, Stefan Lamp, Jonathan Meer, Alfredo Paloyo, Nicolas Salamanca, Christoph M. Schmidt, Stefanie Schurer, Charles Sprenger, Marcus Tamm, Arne Uhlendorff and participants at the the ASSA 2018, 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). We gratefully acknowledge that this work has been partly supported by the Collaborative Research Center *Statistical Modeling of Nonlinear Dynamic Processes* (SFB 823) of the German Research Foundation (DFG), within Project A3, *Dynamic Technology Modeling.* 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. email addresses of authors: 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 the cost of learning arise today, but the expected benefits 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. According to the Economist (Jan 12 2017) they make "balancing learning, working and family life" easier and quickly respond to new skill demands of the labor market with new courses or nanodegrees. Therefore, an emerging literature documents and seeks to explain the massive disengagement in MOOCs. Yet, little is known about how to mitigate it. 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). So far, only few studies aim to tackle these student-related factors with behaviorally-motivated interventions (e.g Martinez, 2014; Patterson, 2016; Kizilcec et al., 2016; Patterson, 2016).[1]

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 where implemented in courses on openHPI and openSAP, two German MOOC platforms operating 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 carried on as usual (Levitt & List, 2009a).

Prompting people to plan has shown beneficial effects in many settings: it increased colonoscopy and mammography uptake (Milkman et al., 2013; Rutter et al., 2006), vaccination (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 two days prior to course start and find no statistically significant effect. Yeomans & Reich (2017) place their planning prompt into a pre-course survey and ask open text questions about MOOC engagement plans and find it increases completion rates by 29%. Our study improves upon these

[1] For an overview of behavioral economics in education see Koch et al. (2015); Lavecchia et al. (2016); Damgaard et al. (2017).

two studies treating all participants irrespective of enrollment timing or survey participation. 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 should be perceived as a course feature. In short, our experimental design should enable us to draw more general conclusions on how planning prompts work in online learning contexts.

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 (2009b).[2] 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. With this design, our paper 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).

Pooling all four courses, we find the planning prompt has no significant effect on the overall certificate rate and course activity. 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 no effects in the other three courses. While finding this positive effect could be a false positive, it may also be plausible that details of course structure matter. Only in a context of infrequent course communication does the planning prompt and its reminder positively influence course activity and completion. Furthermore, the heterogeneity analysis within courses hints at opposite effects for different subgroups. For participants from specific countries, late enrollees and those with a professional interest the planning prompt is beneficial. However, the planning prompt tends to dissuade the participants from the opposite subgroups more.

In light of these results, we conclude that planning prompts are no universal remedy for increasing engagement in online education. Instead, our results imply that planning prompts can be beneficial in courses that lack regular communication and for certain participant groups, but also detrimental for others. 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 the four RCTs. Relying on just 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 and the data. Section 3 discusses potential channels. Section 4 focuses on the results.

[2] According to Levitt & List (2009b), 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.

We first present results from a pooled analysis, then from heterogeneity analyses 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.[3]

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 and the final exam (Renz et al., 2016a). 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.[4]

2.2 Experimental design

We randomly assign participants into control and treatment groups with their first click on the course platform after course start (Figure 1). The treatment is transmitted via pop-ups. Treated participants view a pop-up asking them to determine the concrete time for their next MOOC-session (Figure 2). The pop-up also informs them that they will receive a reminder email shortly before their self-set study time. The treatment therefore is a combination of the planning prompt and the reminder email.

[3] 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 comprised 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.

[4] Please note, to meet the point requirement for the certificate, it is not necessary to hand in all assignments.

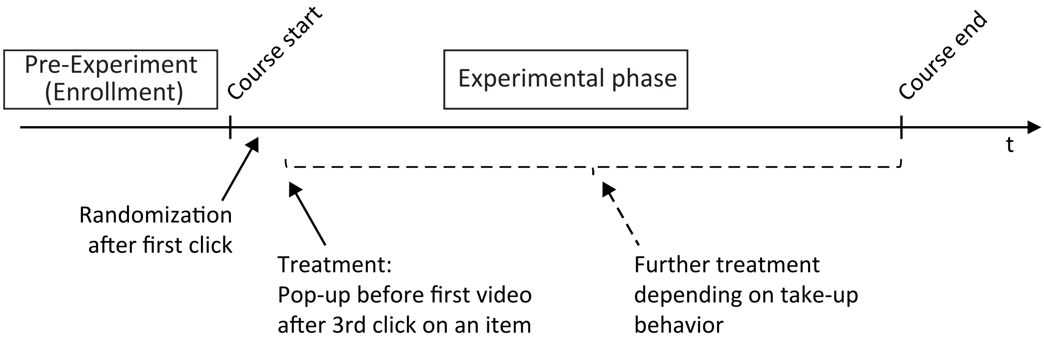
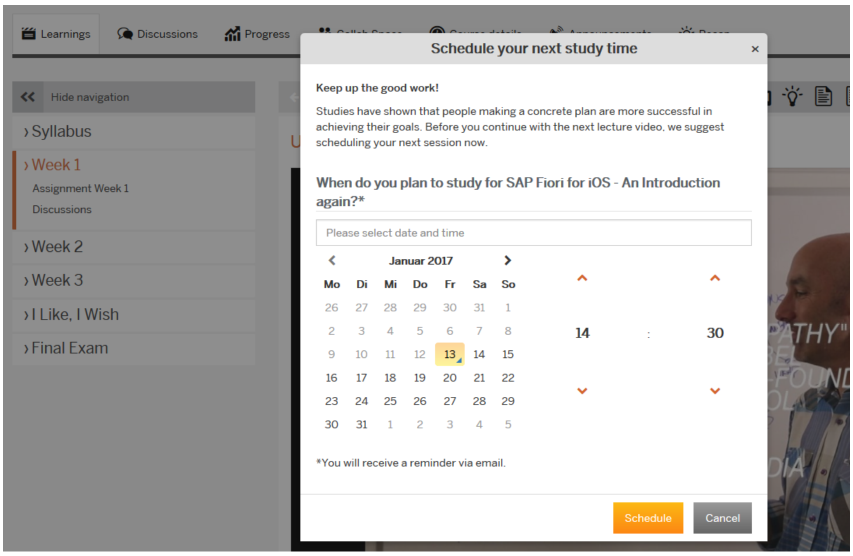


Figure 1: Timeline of the experiment

Take-up of the treatment is optional since the treated participants can use the "cancel"-button 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.[5]

Figure 2: Treatment group pop-up



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. Besides the reminder email of the treatment, both treatment and control groups receive additional reminder emails informing them about the material of a new week, new activities in the course or the start of the final exam.

[5] A session is defined as continuous activity on the MOOC-platform without an interruption lasting longer than 30 minutes.

Figure 3: Control group pop-up

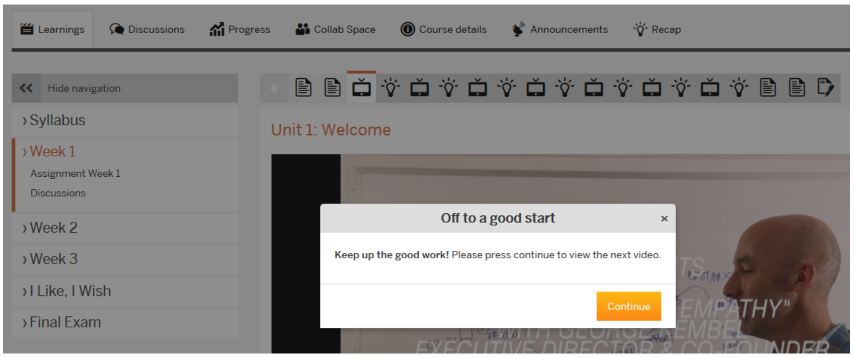


Figure 2 and Figure 3 make clear that the treatment is embedded in the normal user interface and follows the corporate design of the MOOC-providers. This distinguishes our experimental set-up from previous studies. This means all participants in the treatment group are treated and they should perceive the treatment as a course feature. 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. 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 potentially affects 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 they work. This flexibility, they argue, paves the way for time-inconsistent individuals lacking self-control to put off work to later 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 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 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 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, 2016). 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 the planning prompt may not necessarily have positive effects on MOOC completion. For instance, when effort needed for finish the MOOC is not salient, the planning prompt may reduce completion rates. We will return to the discussion of channels, especially the role of time-inconsistency, when analyzing heterogeneous effects.

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, where in Linked there were 2,090 participants; in Web-tech 5,161, in Fiori 2,647, and in Data Science 5,676 participants.

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

Table 1: Summary statistics

	(1) Control	(2) Treated	(3) Difference C & T
Panel A: Pre-determined variables			
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)
Enrolled prior to course start	0.6538	0.6526	0.0012
	(0.4758)	(0.4762)	(0.0076)
Affiliated with SAP	0.0927	0.0859	0.0068
	(0.2900)	(0.2802)	(0.0046)
Panel B: Survey-based or profile-based information			
Answered pre-course survey	0.4541	0.4410	0.0130
	(0.4979)	(0.4966)	(0.0096)
Post-course survey data	0.1837	0.1724	0.0112*
	(0.3872)	(0.3778)	(0.0061)
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)
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)
Course participation due to professional interest	0.6309	0.6436	-0.0127
	(0.4827)	(0.4790)	(0.0138)
Participant is impatient (self-assessed)	0.2733	0.2583	0.0150
	(0.4458)	(0.4378)	(0.0127)
Observations	7704	7870	15574

Note: Panel A shows summary statistics that are collected by the browser. Panel B variables are collected either by the browser or through pre- and post-course surveys or profile mentions. The latter are only available for a non-representative sub-sample. The variables "Course participation due to professional interest", "Course intention" and "Participant is impatient" were collected for Web-tech and Data Science only. Columns 1 and 2 display the sample means with standard deviations in parentheses for the treatment and the control group. Column 3 shows the differences between these two groups and the corresponding standard error. The significance level of a t-test are indicated by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

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. 9% of all participants are professionally affiliated with SAP. 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 year olds and above 50 year olds make up another share of 9% each), and 19% report having a Master's degree or a Phd, while 12% have a Bachelor's degree. However, gender and education information is missing for 60 to 65% of all participants therefore, these numbers may not be very indicative of the overall sample and we will not further analyze them.

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)[6]. A minority (10%) of the sample participates to browse or does not know yet which course outcome they intend (4%). Furthermore, more than one fourth of participants (27%) assesses themselves as impatient.[7]

3.2 Estimation strategy

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

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

where Y_i stands for course completion or intermediary outcomes such as video plays, number of sessions, session duration, number of quizzes submitted, and total points of individual i. T_i indicates the treatment status of individual i. β provides the intention-to-treat effect (ITT) and is the coefficient of interest in this estimation. It captures the causal effect of the treatment T_i on course performance for those participants who were treated – irrespective of whether they actually use the planning tool or not. ε_i captures the remaining idiosyncratic error.

[6]A COP is issued to those participants who have completed at least 50% of the course material.

[7]We asked participants to answer "'Are you generally an impatient person or someone who always shows great patience?"' on a Likert-scale from 0 to 10. Participants who chose the categories 7-10 are classified as impatient.

4 Results

4.1 Pooled analysis

The ITT estimate presented in Table 2 (column 1) for the pooled sample of all four courses reveals that the treatment had no significant effect on the certification rate. The 95%-confidence-interval suggests that potential effect sizes can lie between -7% and 3%. This means any positive effects on overall course completion would be small and only negative effects would be meaningful in magnitude.

Table 2: Certificate completion and intermediate outcomes

	(1) Certificate	(2) Number of sessions	(3) Total duration	(4) Total points	(5) Quizzes submitted	(6) Videos played
Treatment	-0.005	0.158	0.049	-2.115	-0.449	0.265
	(0.007)	(0.432)	(0.195)	(1.464)	(0.463)	(2.066)
Constant	0.277***	18.167***	7.462***	60.419***	23.290***	50.367***
	(0.005)	(0.303)	(0.135)	(1.053)	(0.332)	(1.333)
CI effect sizes	[-7%, 3%]	[-4%, 6%]	[-4%,6%]	[-8%,1%]	[-6%,2%]	[-8%,9%]
Observations	15574	15574	15574	15574	15574	15574

Note: Results are obtained from a OLS regression of the pooled courses. A session is defined as period of interaction with the MOOC platform that is not interrupted for longer than 30 minutes. Total duration is measured in hours. Total points refer to the points that can be earned by submitting assignments or the final exam. Self-test quizzes are not graded. CI effect sizes refer to the 95%-confidence interval (CI). Course fixed effects included. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Similarly the planning prompt does not significantly impact intermediate outcomes which measure course engagement (Table 2, Column 2-6). For one thing, 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. For another, participants have a tendency to earn fewer points and submit less self-test quizzes. The confidence intervals (for points and quizzes) indicate that any meaningful effect sizes are likely to be negative.

Low take-up could be another reason for these non-significant results. Yet, around 30% of all treated participants used the planning tool to schedule their study time at least once.[8] This is a much higher take-up rate than the 13% in Baker et al. (2016), a previous planning prompt intervention in a MOOC context. And 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). Therefore, low take-up is not a plausible factor for the insignificant effects.

[8] We define usage of the planning tool conservatively as scheduling the next study time at least once, and with the scheduled time being at least 2 hours in the future.

Despite our improved experimental design, our results confirm those of previous planning prompt interventions in MOOCs that focused on positively selected samples and where the interventions where not equally-well embedded into the course interface. If anything, Baker et al. (2016) also 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, however, 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.

In sum, the expected positive effect of the planning prompt on course completion could not be detected. If the planning prompt affects MOOC engagement or completion at all, the overall direction is likely to be negative. However, our 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. Previous studies on commitment devices have highlighted adverse effects on subgroups of the sample while at the same time documenting large positive effects on group average (John, 2015). This is why in the following, we explore heterogeneous treatment effects across and within courses.

4.2 Heterogeneity analysis

4.2.1 Heterogeneity across courses

Estimating the ITT separately for each course reveals substantial heterogeneity across courses (Table 3). 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 4). In terms of effect size and significance the number of videos played appears to 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 effect sizes implied by the confidence intervals suggest that the treated participants tend to

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

perform worse than the control group in nearly all engagement indicators (Table 4). In the Fiori course, the coefficients on the number of videos played and total duration even reach statistically significant levels.

While our prior was that a planning prompt would nudge participants positively, these results highlight that meaningful negative effects are not unlikely. For a number of domains, a nascent literature points out that nudge-like interventions such as planning prompts can have disadvantageous effects or even backfire: 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: planning prompts can also potentially dissuade participants from engaging in online courses.

Table 3: Intention-to-treat effects

	(1) Certificate (linked)	(2) Certificate (fiori)	(3) Certificate (datas)	(4) Certificate (webtech)
Treated	0.034*	-0.018	-0.013	-0.006
	(0.017)	(0.019)	(0.012)	(0.012)
Constant	0.178***	0.355***	0.303***	0.250***
	(0.012)	(0.013)	(0.009)	(0.009)
CI effect sizes	[0;38%]	[-16%; 5%]	[-12%; 3%]	[-12%; 7%]
Observations	2090	2647	5676	5161

Note: Results are obtained from a reduced-from-regression for each course separately. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

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 the course structure. The Linked course does not differ much in most observable course characteristics, such as duration, video time, number of quizzes (see mean values and Z-Scores in Table 5). Yet, it deviates from the mean of all courses in the number of emails alerts per week. Due to a programming error only half of the usual number of the emails informing about new content were sent out successfully. 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 do not appear beneficial in contexts with frequent emails.

Table 4: Effect of planning prompt on intermediary outcomes

	(1) Number of sessions	(2) Total duration	(3) Total points	(4) Quizzes submitted	(5) Videos played
Panel A: Linked					
Treated	1.139	0.744	2.867	1.868	11.443*
	(1.196)	(0.546)	(2.369)	(1.278)	(5.892)
Constant	18.210***	7.473***	30.987***	20.499***	46.519***
	(0.813)	(0.350)	(1.622)	(0.877)	(2.811)
CI effect sizes	[-7%;19%]	[-4%;24%]	[-6%;24%]	[-3%;21%]	[0%;49%]
N	2091	2091	2090	2091	2091
Panel B: Fiori					
Treated	-0.330	-0.240*	-2.899	-0.897	-4.335**
	(0.340)	(0.135)	(2.792)	(0.555)	(1.851)
Constant	8.882***	3.173***	62.909***	15.153***	25.728***
	(0.243)	(0.106)	(1.980)	(0.399)	(1.478)
CI effect sizes	[-11%;4%]	[-16%;1%]	[-13%;4%]	[-13%;1%]	[-31%;-3%]
N	2647	2647	2647	2647	2647
Panel C: Web-tech					
Treated	0.088	0.060	-0.816	-1.139	-4.367
	(1.022)	(0.467)	(1.734)	(1.025)	(3.759)
Constant	26.954***	12.001***	43.178***	33.222***	69.518***
	(0.726)	(0.330)	(1.247)	(0.749)	(2.796)
CI effect sizes	[-7%;8%]	[-7%;8%]	[-10%;6%]	[-9%;3%]	[-17%;4%]
N	5161	5161	5161	5161	5161
Panel D: Data Science					
Treated	-0.167	-0.215	-4.655	-0.773	1.972
	(0.484)	(0.194)	(3.227)	(0.603)	(3.835)
Constant	14.624***	5.399***	85.724***	19.244***	46.180***
	(0.341)	(0.137)	(2.325)	(0.428)	(2.291)
CI effect sizes	[-8%;5%]	[-11%;3%]	[-13%;2%]	[-10%;2%]	[-12%;21%]
N	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. A session is defined as period of interaction with the MOOC platform that is not interrupted for longer than 30 minutes. Total duration is measured in hours. Total points refer to the points that can be earned by submitting assignments or the final exam. Self-test quizzes are not graded. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 5: 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: Z-Scores in parenthesis.

4.2.2 Heterogeneity within courses

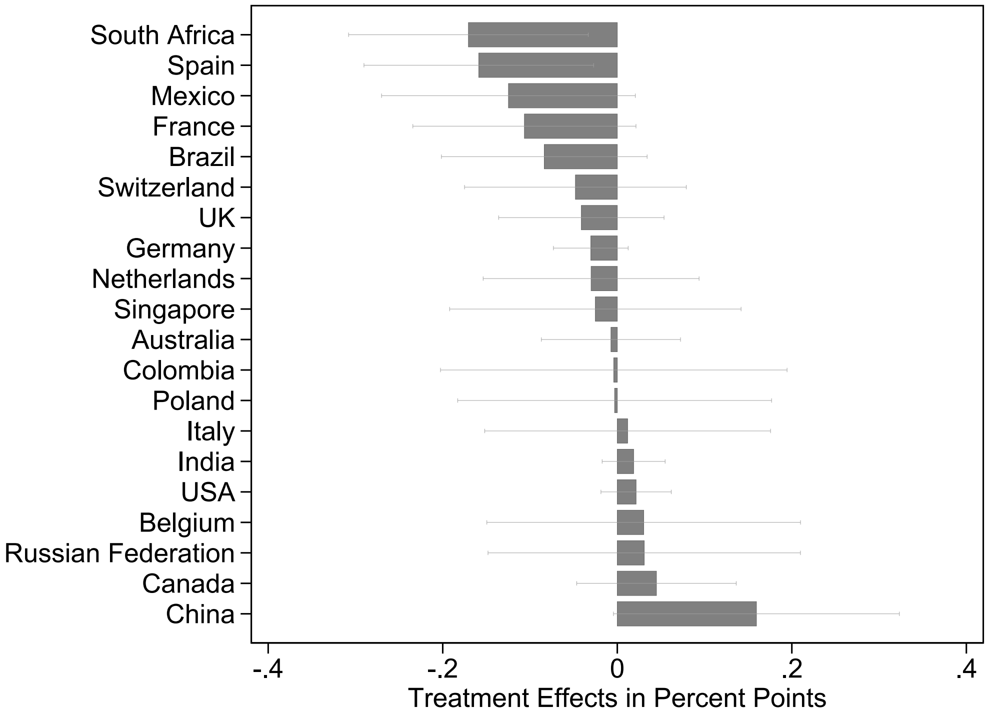
The overall insignificant effect of the planning prompt on course completion may also be brought about by opposing effects on different subgroups of participants with in courses. We investigate characteristics which the literature has identified as important, such as enrollment timing (Banerjee & Duflo, 2014), country of origin (Kizilcec & Cohen, 2017), interest and intention (Reich, 2014), and proxies for time-inconsistency (impatience and procrastination) which we hypothesized would be mitigated by planning prompts.[10] These characteristics are either measured by the browser and observable for nearly all participants (Table 6) or by a pre- or post-course survey implying that they are only observable for a selective subsample (Table 7). We exclude the Linked course from this analysis due the slightly different course structure. Our explorative heterogeneity analysis finds tentative evidence for such offsetting effects.

There are substantial heterogeneities by country of origin. Figure 4 illustrates the range of treatment effects by country: while the planning prompt appears to negatively effect South African participants, decreasing their completion rate by 17.5 percentage points compared to the control group, the planning prompt tends to encourage completion for Chinese participants, increasing their completion rate by 15.6 percentage points. Kizilcec & Cohen (2017) highlight that the effectiveness of behaviorally motivated interventions in MOOCs may depend on cultural contexts. Despite most point estimates being insignificant, they are statistically significant different from each other, hence our results support this view.

Inspecting this heterogeneity further, we correlate the treatment effects of these 20 countries with the culture dimensions power distance, masculinity, individualism and uncertainty avoidance, which Hofstede (1986) suggests are relevant in education contexts, and patience (Falk et al., 2018), which to some extend captures time inconsistent preferences. In contrast to Kizilcec & Cohen (2017), these plots suggest that while culture and patience may play

[10]Please note, we only specified a subset of these covariates in our pre-analysis plan and therefore refer to most of this analysis as explorative.

Figure 4: Treatment effects by country

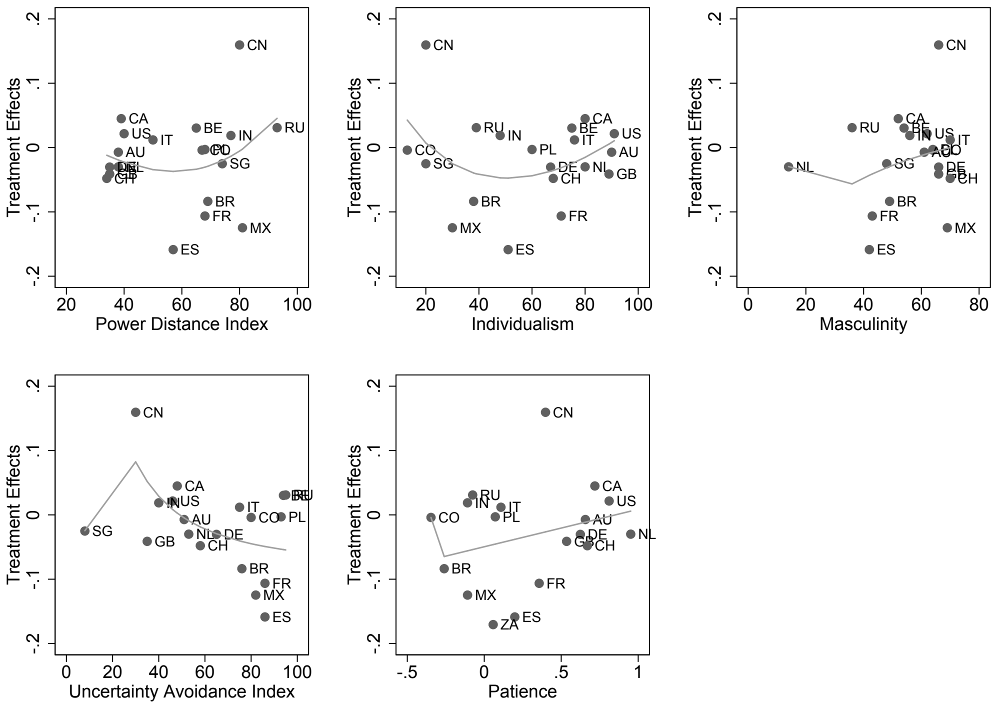


Notes: The figure displays the average difference in certification rates between treatment and control group by country in percentage points. We only use the openSAP-courses Fiori and Data Science which have an internationally diverse participant pool. Only countries with more than 30 participants in treatment or control group are included in the graph. 90%-confidence intervals are shown. We control for course fixed effects.

a role the associations do not appear to be linear. For instance, the patience scatter plot suggests three different country clusters when excluding outliers like China.

There is also heterogeneity by professional interest, which we approximate by the email affiliation used for enrollment (Panel D in Table 6). 9% of participants of the Fiori and Data Science course enroll using an SAP email address. This means they are professionally associated with the company in some way, which suggests that they have a strong professional interest in the course. In general, SAP-affiliates perform significantly worse than non-affiliates; however, they are positively influenced by the planning prompt. Being prompted to plan increases their certification rate by 2.6%-points (8%) compared to the control group's 31% control rate. This difference between affiliation and non-affiliation is statistically significant, however, the marginal effect of treatment on the affiliated does not reach statistically significant values. A professional interest in the course may go hand-in-hand with a professional context in which scheduling meetings or blocking time is natural. In such a context the planning prompt can be beneficial.

Figure 5: Correlating Treatment Effects with Culture Indicators and Preferences



Notes: The figure summarizes the correlation of treatment effects by country and cultural dimensions by country as they were defined by Hofstede et al. (2010) and global preferences elicited by Falk et al. (2018). These data are available online: https://geerthofstede.com/research-and-vsm/dimension-data-matrix/ (accessed: March 14, 2018) and https://www.briq-institute.org/global-preferences/downloads (accessed: July 13, 2018) We use the openSAP-courses Fiori and Data Science which have an internationally diverse participant pool. Only countries with more than 30 participants in treatment or control group are included in the graph. We control for course fixed effects. Fractional-polynomial prediction displayed as line in each plot.

Table 6: Heterogeneities: non-survey-based characteristics

	Certificate		Certificate
A: Professional interest		**B: Enrollment timing**	
Treated	-0.023**	Treated	-0.057*
	(0.011)		(0.03)
Affiliated with SAP	-0.050***	Late	-0.092***
	(0.019)		(0.026)
Treated x Affiliated with SAP	0.049*	Treated x late	0.063*
	(0.027)		(0.036)
Constant	0.313***	Constant	0.399***
	(0.009)		(0.024)
Observations	8323	Observations	2833

Notes: Table shows the heterogeneities of characteristics not elicited by a survey. The number of observations differ depending on the courses pooled. Panel A, country, use data from the two openSAP courses: Data Science and Fiori, controlling for missing country information. Panel B uses data from the Fiori, Data Science and the web-tech course in a 5 day window around the course starting dates. Panel C uses data from the Fiori, Data Science and the Web-tech course. Panel D, professional interest approximated by SAP affiliation. Course fixed effects are included. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Moreover, we find heterogeneities with regard to self-organization skills. Banerjee & Duflo (2014) suggest that participants who enroll before the MOOC starts are better in self-organization and therefore, perform better in the course. This also holds true in our study. Participants who enroll within five days after the course start have a 9.2 percentage point lower probability of earning a certificate than those who enrolled in the five days before course start. Yet, overall the planning prompt does not make late enrollees more persistent (Panel B, Table 6). While late enrollees react significantly different from early enrollees the overall treatment effect is small and insignificant. The planning prompt, however, does seem to repel early enrollees. They are 5.7 percentage points (14%) less likely to earn a certificate. This discouraging effect of the planning prompt may be the result of annoying those who are well-organized. In one course, anecdotal evidence supports this explanation. One participant commented "If I see a need to schedule the time to study, I would set a reminder on my own."

Furthermore, we test whether making course effort more salient early in the course induces quicker dropout as some theories suggest. Table 8 shows that he treatment does not significantly affect the probability to being active in the course for more than one week. In addition to insignificance the point estimate of 0.007 percentage points is very small compared to the 60% of control group participants, who are still present after their first week. Hence, this suggests that the increased saliency of effort does not expedite drop-out and more general that the timing of dropout does not seem to drive the tendency to disengage from the course.

Table 7: Heterogeneities: survey-based characteristics

Certificate **A: Pre-Course Survey**		Certificate **B: Course intention**		Certificate **C: Impatience**	
Treated	-0.000 (0.011)	Treated	-0.039 (0.027)	Treated	-0.016 (0.016)
Pre-Survey	0.272*** (0.012)	Earn ROA	0.163*** (0.023)	Impatient	-0.018 (0.022)
Treated x Survey	-0.0132 (0.016)	Browse	-0.130*** (0.036)	Treated x impatient	0.011 (0.031)
		Don't know	0.062 (0.055)		
		Treated x Earn ROA	0.04 (0.032)		
		Treated x Browse	0.027 (0.05)		
		Treated x Don't know	-0.031 (0.075)		
Constant	0.155*** (0.008)	Constant	0.480*** (0.022)	Constant	0.600*** (0.014)
Observations	10837	Observations	4849	Observations	4849

Certificate **D: Post-Course Survey**		Points achieved **E: Procrastination**	
Treated	0.007 (0.007)	Treated	-2.060 (4.560)
Post-Survey	0.782*** (0.012)	Occasional	1.125 (4.379)
Treated x Post-Survey	-0.0128 (0.018)	Chronic	-3.569 (5.010)
		Treated x occasional	3.664 (6.384)
		Treated x chronic	4.870 (7.240)
Constant	0.135*** (0005)	Constant	269.665*** (3.456)
Observations	10697	Observations	1735

Notes: Table shows the heterogeneities of characteristics elicited by a survey. All panels use data from the Web-tech and Data Science course. The number of observations differ depending on the survey response. Course fixed effects are included. Panel B also controls for missing intention information of those who otherwise responded to the survey. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 8: Effect of planning prompt duration

	(1) Course duration longer than a week
Treated	-0.007
	(0.008)
Constant	0.595***
	(0.008)
Observations	13484

Notes: This table presents the results from a regression of the treatment on an duration variable, indicating whether the participant is present in the course for more than a week. Course fixed effects included. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Intentions for taking a MOOC may be very different (Reich, 2014). With this in mind, disengagement from a MOOC is particularly worrisome for those who actually intend to earning a certificate. In the the Web-tech and Data Science courses we asked participants about their intentions in a pre-course survey. 57% and 33% of all participants in Web-tech and Data Science answered the pre-course survey respectively.[11] 61% state the intention to earn a Record of Achievement (ROA), a quarter indicates wanting to earn a Certificate of Participation (COP), and 10% intend to browse. Our results confirm that course intentions are positively correlated with certificate rates (Panel B, Table 7). Despite intentions do not translating into actions 100%, the planning prompt does not affect participants with different course intentions significantly differently. The point estimates suggest that those with an intention to earn a ROA tend to be positively affected, while those who intend to browse are unaffected and those who would like to earn a COP are discouraged.

In addition, we examine whether the planning prompt helps mitigate time-inconsistent preferences and self-control problems. First, we elicit time preferences with a survey question from the German Socioeconomic Panel (SOEP) (Vischer et al., 2013). 27% of participants assess themselves as impatient. Contrary to our prior, this survey-based impatience does not appear to be a driving factor in completing a MOOC (Panel B, Table 7) because the point

[11] The question was asked slightly differently in both courses. In the Web-tech course the question read: "What is your primary goal and motivation in taking this course?" and the answer categories were: "a) engaging with the course material by reading the material and doing the tests to earn a certificate at the end, b) learning the course topics by watching videos c) None of the above. I would like to browse the course." For Data Science the question read: "Users have different intentions for taking a course. Which of the following best applies to you?" Answer categories were a) "I am here to browse the material, but not planning on earning a Record of Achievement or Confirmation of Participation. b) I am here to view the material and planning to obtain a Confirmation of participation. c) Here to work through the material and planning on collecting enough points to earn a Record of Achievement. d) I have not decided how I would like to engage with the course material." When pooling we code categories a (Web-tech) and c (Data Science) to the category "Earn a Record Achievement", b (Web-tech) and b (Data Science) to "Earning a Confirmation of Participation" and c (Web-tech) and a (Data Science) to "Browse course".

estimate on "impatient" is small and insignificant. Therefore, it is not surprising that the planning prompt does not significantly affect impatient differently than patient participants.

Second, we elicit procrastination tendencies of participants in the post-course survey of Data Science and Web-tech.[12] 11% and 15.7% replied to the post-course survey in the Web-tech and Data Science course respectively.37% indicate to procrastinate infrequently, 39% occasionally procrastinate, and about one quarter of participants chronically procrastinate. While all point estimates are statistically insignificant the signs and magnitudes suggest plausible tendencies (Panel E, Table 7): treated infrequent procrastinators may earn less points than their control group counterparts as a result of the planning prompt, whereas occasional and chronic procrastinators are more likely to earn a few more points. This suggests that the planning prompt may have helped lessen the impact of procrastination. The results for patience and procrastination do not allow for clear-cut conclusions whether time-inconsistent preferences are an important obstacle for completing a MOOC and whether the planning prompt can serve as a remedy.

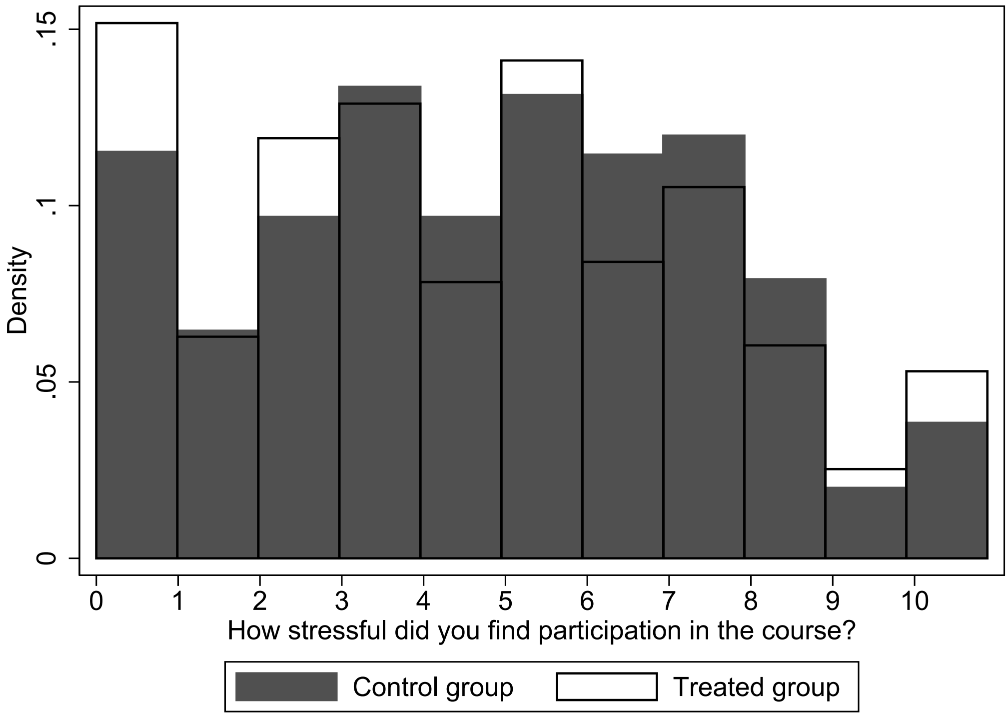
In sum, we find tendencies for small positive heterogeneous effects of the planning prompt for some subgroups – namely for participants from some countries India, the US, Canada, and China, for late enrollees and professionally interested. We also revealed negative effects for the other groups, such as South African, Spanish, French, and Brazilian participants, participants without strong professional interest and early enrollees. The dissuading effects are often larger in size and affect larger participant groups than the encouraging effects. Hence, the overall effect is attenuated towards zero. This corroborates the impression that overall effects of meaningful size are likely to be negative; but some groups are unaffected or react slightly positively to the planning prompt. Furthermore, we find indicative evidence that increased salience of effort may not play a role, since the planning prompt does not make participants drop out sooner.

5 Potential side Effects of the Planning Prompt

Previous literature shows that using nudge-like interventions may also have disadvantageous effects on overall welfare even when there are desired positive effects on the main outcome variable (c.f. Damgaard & Gravert, 2018; Allcott & Kessler). 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. Consequently, the 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

[12]Participants were asked to rate on a 0-10 Likert scale "Do you generally procrastinate?". We classify all participants with answers between 1-3 as infrequent procrastinators, with 4-6 as occasional procrastinators and 7-10 as chronic procrastinators.

Figure 6: Stress level by treatment for all courses



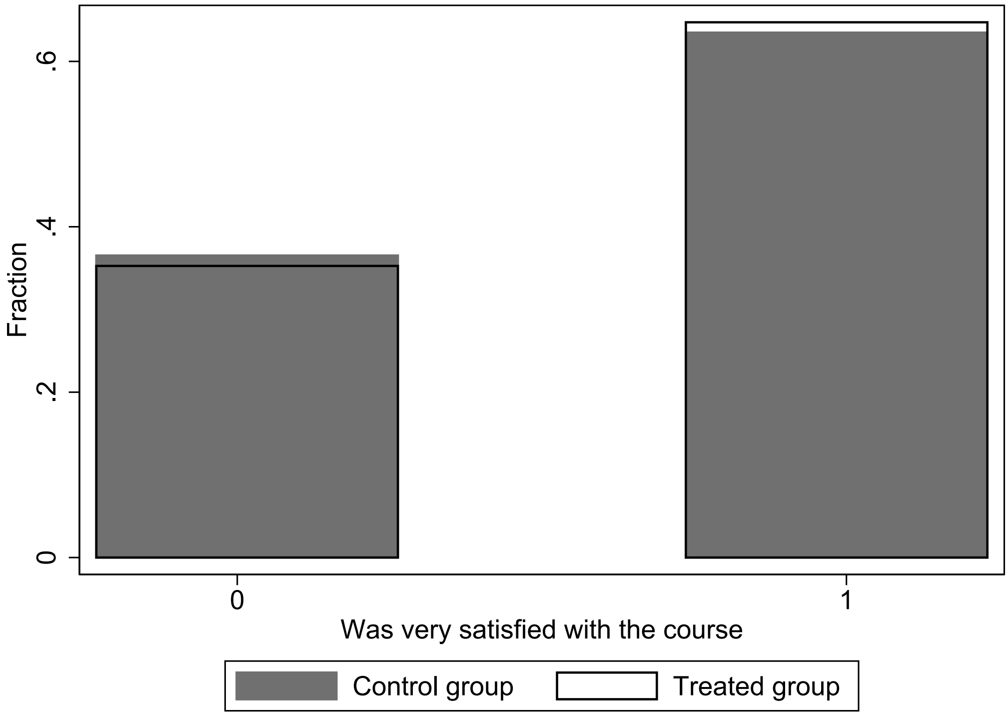
survey: these were 282 participants (13.49%) in Linked, 984 (18.86%) in Web-tech, 612 (23.12%) in Fiori, and 893 (17.20%) in Data Science. However, this caveat applies to the treatment group as well as to the control group making the comparison viable.

Course participants in the treatment group less often report high stress levels 7-10 than the control group but tend to report being less stressed (Figure 6). Furthermore, we check whether overall course satisfaction was affected by the planning prompt. Figure 7 shows no signs of this. The treatment and control groups in both courses seem equally satisfied with the course. We therefore conclude that the planning prompt did not bring about a worse course experience for the treated who complete the course.

6 Conclusion

This paper examines whether prompting participants to plan 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 no overall effect on course completion or course engagement. Yet, there are substantial heterogeneities across and within courses. Apparently small differences in course structure matter with respect to the planning prompt's impact: in one course, treated participants have a 19% higher probability of completing the course with a certificate, but the other three courses show no significant effects. While the positive

Figure 7: Satisfaction level by treatment for all courses



Notes: Very satisfied refers to indicating satisfaction 4 and 5 on a 5-point Likert scale.

effect could be a false positive, we also discuss plausible reasons why specific details of the course structure might make a decisive difference. Furthermore, we find heterogeneous effects with respect to participant characteristics like country of origin, professional interest, and self-organization. While subgroups with respect to these characteristics are significantly positively influenced by the planning prompt, the performance of the others is negatively affected by the planning prompt. Yet, the planning prompt did not result in worse course experience in terms of course satisfaction and stress levels.

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 and target all participants than 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 small nudges are powerful enough to help overcome barriers to following through with education intentions like present bias, overconfidence, inattention and lack of salience (Patterson, 2016; Bisin & Hyndman, 2014; Burger et al., 2011). Instead, the heterogeneity analyses implies that the 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 one detail like the frequency of email communication can influence signs and effect sizes substantially. Instead of extrapolating findings from one study context to another, it is useful to replicate (even in the first study) especially if the costs of doing so are low, as 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. (????). 2018 forthcoming. The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics*.

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

Banerjee, A. V., & Duflo, E. (2014). (Dis) organization and Success in an Economics MOOC. *The American economic review*, *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.

Bisin, A., & Hyndman, K. (2014). Present-Bias, Procrastination and Deadlines in a Field Experiment. *NBER working paper series*, *19874*.

Burger, N., Charness, G., & Lynham, J. (2011). Field and online experiments on self-control. *Journal of Economic Behavior & Organization*, *77*(3), 393–404.

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

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., Nielsen, H. S., et al. (2017). Nudging in education: A survey. Tech. rep.

Dehejia, R., Pop-Eleches, C., & Samii, C. (2015). From Local to Global: External Validity in a Fertility Natural Experiment. *NBER Working Paper Series*, (21459).

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`

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences*. *The Quarterly Journal of Economics*, (p. qjy013).

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)`

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, *10*(3), 301–320.

Hofstede, G., Hofstede, G., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*. New York: McGraw-Hill Education, third edition ed.

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., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences*, *114*(17), 4348–4353.

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 (2016) 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. (2009a). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

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

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.

Patterson, R. W. (2016). Can Behavioral Tools Improve Online Student Outcomes? Experimental Evidence from a Massive Open Online Course.
URL https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxyaWNocGF0dGVyc29uY29ybmVsbHxneDoxZDA0MzM2YjRlMjM2YTg0

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. (2016a). openSAP: Evaluating xMOOC usage and challenges for scalable and open enterprise education. *International Journal of Advanced Corporate Learning*, *9*(2), 34–39.

Renz, J., Schwerer, F., & Meinel, C. (2016b). openSAP: Evaluating xMOOC usage and challenges for scalable and open enterprise education. In *The International Conference on E-Learning in the Workplace 2016*.
URL `www.icelw.org`

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.

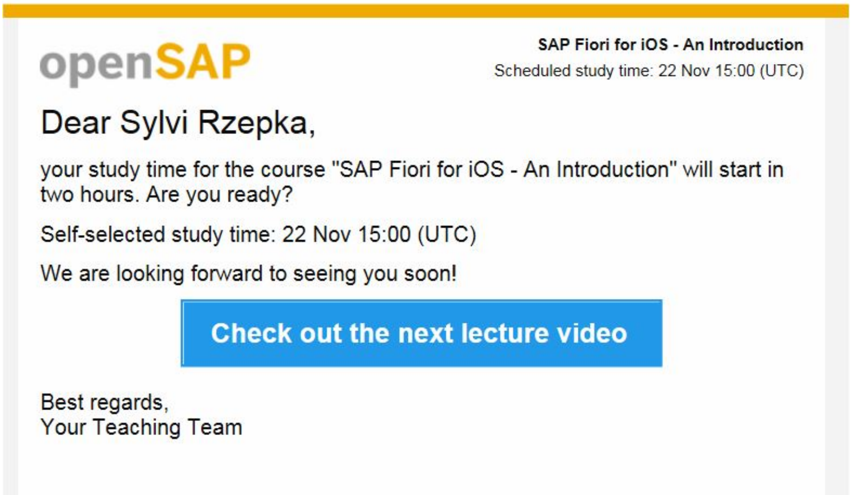
Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, *120*(2), 142–145.

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

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



openSAP

SAP Fiori for iOS - An Introduction
Scheduled study time: 22 Nov 15:00 (UTC)

Dear Sylvi Rzepka,

your study time for the course "SAP Fiori for iOS - An Introduction" will start in two hours. Are you ready?

Self-selected study time: 22 Nov 15:00 (UTC)

We are looking forward to seeing you soon!

Check out the next lecture video

Best regards,
Your Teaching Team