Rabbits and Study Habits: A Field Experiment on Pacesetters and Student Effort*

Gert-Jan Romensen†

University of Groningen

July 31, 2020

Abstract

Many students perform below their academic potential due to a lack of study time. Targeting study time and evaluating the effects has thus far been challenging because of poor measurement of effort and the absence of information on students' initial study plans. In a field experiment with 573 university students, I elicit detailed weekly study goals and plans from students and use these to construct individualized pacesetters (*rabbits*). Pacesetters are moving reference points that visualize the preferred study pace of the present self by moving exactly according to the initial study plan. I build a new educational technology to measure effort and to display the pacesetters to students in real time. Falling behind the pacesetter confronts the student with his or her procrastination. I find that students have more ambitious study goals when they set a pacesetter, but subsequently tend to fall behind quickly and are less likely to reach their goals. The pacesetter has no impact on study time, nor does it improve learning outcomes. I discuss treatment heterogeneity based on: gender, ability, and procrastination.

JEL classification: I21, D91

Keywords: student effort, learning inputs, pacesetters, procrastination

*I thank Josse Delfgaauw, Marco Haan, Jonathan Meer, Uros Petronijevic, Adriaan Soetevent, Dinand Webbink and seminar participants at AFE 2019, ESA European Meeting 2019, IAREP/SABE 2019, KVS New Paper Sessions 2019, NCBEE 2019, Ca' Foscari University of Venice, and University of Groningen for their valuable comments. Further thanks to Wim-Paul Hiddink and Mark van Oldeniel for their excellent research assistance. The educational technology used in this paper is co-developed with Tonny Romensen. My deepest gratitude to him for enabling this project. Views and opinions expressed in this paper as well as all remaining errors are solely those of the author. Financial support by GrEELab experiments is gratefully acknowledged. This study is registered in the AEA RCT Registry and the unique identifying number is: "AEARCTR-0003356".

†University of Groningen, Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen, The Netherlands, g.j.romensen@rug.nl.

1 Introduction

A student's own study effort is a core input in the education production function and is the one input that is under direct control of the student. Various studies stress and show that a lack of study effort has detrimental effects on educational achievements (Metcalfe, Burgess and Proud 2019, Stinebrickner and Stinebrickner 2008, Angrist, Lang and Oreopoulos 2009). Observed declines in students' study effort are therefore especially worrisome (Oreopoulos, Patterson, Petronijevic and Pope 2018). Between 1961 and 2004, the average weekly study time of college students has decreased by about a half (Babcock and Marks 2011). In light of OECD figures showing that delayed graduation is common and that about 31% of the students drop out of tertiary education (OECD 2008), the problem of study effort decline thus seems acute and in urgent need of attention from researchers.

How can students be motivated to study more? Policies that can raise student effort may have great promise in improving various educational outcomes (Lavecchia, Liu and Oreopoulos 2016). Implementing such policies, however, is widely considered a challenge. It requires detailed data on effort and students' initial study plans in order to provide timely feedback on individual progress. Such data are often absent in existing research. Furthermore, students tend to have self-control problems in how effort is allocated over time (Augenblick, Niederle and Sprenger 2015, Wong 2008). Procrastination is widespread and estimated to affect up to 95% of tertiary students (Steel 2007). These issues have stalled progress on designing and implementing policies that can meaningfully alter effort.

In this chapter, I aim to overcome these issues by developing and leveraging a new educational technology, called the *MicroApp* (an online learning platform), to meticulously measure student effort and to administer personalized planning interventions to students in real time. Specifically, in a field experiment I examine the effect on student effort when prior study commitments by students are presented as real-time virtual pacesetters (*rabbits*). In sports, a rabbit is the colloquial term used for a runner who is instructed to run at the exact pace needed to finish at a certain time. In education, pacesetters can provide a succinct and salient reference point that visualizes the preferred study pace envisioned by the present self. The distant study goal is transformed into a continuous and easy-to-interpret task that activates students directly: stay close to your own pacesetter if you want to reach your initial study goal.

As the pacesetter is, by design, not prone to self-control problems, falling behind immediately makes salient to students that they are reneging on earlier commitments. Under the assumption that students exhibit reference-dependent preferences, loss aversion makes falling behind psychologically painful and motivates the student to catch up and exert effort. The pacesetter serves as a commitment whereby future selves are induced to follow the pacesetter and, in doing so, implement the initial study plan.

I use the phase-in of the MicroApp at a Dutch university to implement a field experiment that identifies the causal effects of the pacesetter on effort and learning outcomes. The experiment covers a large introductory course in microeconomics with a total enrollment of 573 students. The MicroApp is a self-developed online learning platform that is specifically created for this course. The platform contains course recaps and ample practice material with feedback. Students can use the platform during self-study and select at any time the course topics they want to work on. The MicroApp is offered free-of-charge to students and is a non-mandatory component of the course. This way, I can treat platform usage as a measure of pure study time, and not as a measure of actions undertaken to directly satisfy some course criterium. Student activity on the platform is minutely recorded and registered in individual user logs.

In the experiment, students are asked each week through in-class surveys to formulate a study goal and plan regarding MicroApp usage in the upcoming week. I ask students in the *control* group only to formulate a study goal. Their MicroApp version will show how much of that goal is achieved at every moment. I ask students in the *planner* group for a detailed study plan in terms of MicroApp usage: when and how long do they plan to use the platform? Their MicroApp version allows them to check back on the study plan they submitted. Students in the *rabbit* group also submit a study plan, but their MicroApp has the pacesetter feature: it visualizes how much time they spent so far this week, and how much time they were planning to have spent. In other words, they observe in real time whether they are ahead or behind their own pacesetter.

By comparing usage of the MicroApp across all three conditions, a detailed picture is created of student effort and the distribution of this effort over time. Matching these data with other survey information and administrative databases, I can check how usage affects other learning inputs (e.g., attending lectures) and learning outcomes (e.g., final exam score).

The results are as follows. The pacesetter induces students to set more ambitious study goals, amounting to about two additional study hours per week. Similar effects are observed, albeit smaller, for students in the planner condition. Interestingly, students in both treatments are less likely to complete their study goals on time, partly as a result of having set too ambitious study goals. I find no evidence that the treatments increase study time on the MicroApp, nor do I observe effort spillovers to other learning inputs or other courses. Further analyses reveal no treatment impact on various learning outcomes. Treatment heterogeneity based on gender, ability, and procrastination is considered in detail but found to be only limited in terms of effort, goals, and learning outcomes. The null results are in line with research on low-cost interventions that target study time (Oreopoulos et al. 2018).

This study is one of the first attempts to use educational technologies as a means to target behaviors that have a negative impact on academic performance. I show the potential of using these technologies to measure student effort, to test interventions rigorously by implementing field experiments, and to gather insights on how students learn. The digitalization of education sets a new research agenda and unveils many research opportunities. It will require from researchers and educators alike to think clearly about how students learn in an increasingly digital world.

This chapter proceeds as follows. Section 2 discusses related literature. The research question and corresponding hypotheses are covered in Section 3. In Section 4, I elaborate on the experimental design and setting. The empirical evaluation is presented in Section 5. Section 6 provides a brief further discussion of the results and details the implications. The conclusions can be found in Section 7.

2 Related literature

Studying is a classic investment decision, as upfront costs need to be incurred first before benefits can be enjoyed. Whether students can properly execute such a decision is subject to much research (Lavecchia et al. 2016, Koch, Nafziger and Nielsen 2015). Central to the discussion is evidence that many students display self-control problems with regard to how effort is allocated over time (Augenblick et al. 2015, Wong 2008). They tend to have difficulties following through on earlier commitments and as a result may provide insufficient effort to reach initial goals. Many students perform poorly in educational programs due to a simple lack of preparation (Oreopoulos et al. 2018, Angrist et al. 2009).

Students are at least partially aware of their own self-control problems (Wong 2008), and they may therefore seek and benefit from commitment devices (Bryan, Karlan and Nelson 2010, O'Donoghue and Rabin 1999). In education, a commonly used commitment device is imposing a deadline on task completion. For example, evenly spread deadlines may alleviate self-control problems by inducing students to start on a task earlier. Evidence for this is mixed, however, with some studies reporting positive effects (Ariely and Wertenbroch 2002) and others reporting no effects (Burger, Charness and Lynham 2011). This suggests that deadlines still offer too much leeway for intertemporal effort substitution. Other settings provide further suggestive evidence of this: people bunch their effort right before the deadline when filing taxes (Martinez, Meier and Sprenger 2017) and paying traffic fines (Heffetz, O'Donoghue and Schneider 2016).

Recent studies propose goal bracketing, an internal commitment device, to address the issue of study effort being postponed. With goal bracketing, individuals endogenously set subgoals and adopt mental accounts to evaluate their performance in a broad or narrow performance account (Koch and Nafziger 2016, Koch and Nafziger 2017, Hsiaw 2015). In theory, narrow goals (e.g., daily instead of weekly study goals) can be particularly effective for procrastinating students because they rule out behavior in which lack of effort today can be compensated for with more effort in the future. The trade-off is that it comes at the cost of reduced flexibility. Furthermore, there may be implementation issues. Facilitating goal bracketing requires high-frequency data on effort provision at the student level in order to verify whether goals are met. Also, if narrower goals insulate students better from self-control problems, there may be a risk of feedback with excessive goal bracketing. Providing progress updates on too many goals may overwhelm students and dilute the motivational effects of goal-setting.

In most of the existing research on ways of increasing student effort, actual effort by students is unobserved due to data limitations. Output measures (e.g., exam performance) are used instead to approximate the provided effort. This leads to two issues. First, learning outputs do not reveal how student effort was distributed over time, and to what extent this realized distribution matches initial study plans. Second, ability may confound results when learning outputs are used as proxies for effort levels. Disentangling the effects of ability and effort is crucial for the design of educational interventions (Levitt, List, Neckermann and Sadoff 2016). When the goal is to motivate students, it may be promising to divert attention from output and to directly target and incentivize learning inputs instead (Clark, Gill, Prowse and Rush 2017, Lavecchia et al. 2016).

Educational technologies show considerable promise in tailoring interventions that target behaviors with a negative impact on academic performance (Escueta, Nickow, Oreopoulos and Quan forthcoming). Student effort and performance can be recorded at low cost by these technologies. This makes it possible to target interventions at learning inputs, and to assess the impact with the help of user data. Technologies also facilitate new ways of visualizing feedback, which makes it easier for students to digest greater amounts of information on progress. Furthermore, by asking students beforehand what their study plans are, technologies can precisely map to what extent actual effort over time matches initial study plans. To date, however, there is little research available that has explored how and which behavioral interventions can be integrated in educational technologies in order to examine and address self-control problems among students.

This study aims to fill this gap. I elicit initial study plans from students and use a novel educational technology to measure how actual effort compares with these plans. I then build on existing goal-setting research and propose an individualized self-study pacesetter (a moving reference point) to keep students on track. The pacesetter preserves flexibility (students can catch up), but still triggers the motivating effect of loss aversion when students fall behind. It has the additional benefit of condensing a large amount of information into a single focus point. A field experiment is implemented to identify the treatment effects of the pacesetter.

Conceptually, a useful way of thinking about the pacesetter feature is to see it as a visualization of the preferred study pace envisioned by the present self. A common view in the literature on intertemporal decision making is that there are conflicts between different selves (Hershfield 2011, Frederick 2003), as in models where an individual is both a long-term planner and a short-sighted doer (Benabou and Pycia 2002, Thaler and Shefrin 1981).[1] Such conflicts may be resolved when the different selves can better connect to each other. Technology may provide the means to achieve this. Hershfield et

[1] In their discussion of long-term educational decisions, Lavecchia et al. (2016) also propose a dual-system approach, with one forward-looking system and one that is not. Already in 1759, Adam Smith wrote that behavior is determined by the struggle between the "passions" and the "impartial spectator" (Ashraf, Camerer and Loewenstein 2005).

al. (2011), for example, use ageing software and find that people save more for the future when presented age-progressed renderings of themselves. In this chapter, future selves are triggered to connect with the present self through technology-enabled virtual pacesetters.

Clark et al. (2017) introduce a model of task-based goal-setting that addresses the conflict between different selves in the context of study effort. In their model, a student "planner" first sets a goal $g \geq 0$ for the number of study tasks to complete $a \geq 0$. Performance on the exam increases in $a : f'(a) > 0$. The loss-averse student "actor" then chooses the level of a and suffers goal disutility when a falls short of $g : -\lambda \max\{g - a, 0\}$, where λ captures the strength of goal disutility from loss aversion. Finally, the utility of the student "beneficiary" increases in exam performance. It follows that for a student actor exhibiting quasi-hyperbolic discounting (Laibson 1997), the utility function becomes: $\mu_{act}(a|g) = \beta\delta f(a) - [\lambda\{g - a, 0\} + C(a)]$, where $C(a)$ is the cost of effort, $\beta \in (0, 1)$ is the present bias, and $\delta \in (0, 1]$ is the standard discount factor. Crucially, unlike performance-based goals, the utility function shows that task-based goals result in immediate disutility for the student actor when $a < g$. Temporal distance $(\beta\delta)$ that would dampen the motivational effect of the goal is thereby eliminated.[2] The pacesetter feature does exactly that and shows how temporal distance can be removed by letting students "race" against their own effort-based goals.

3 Research question and hypotheses

This chapter aims to address the following question: do students exert higher study effort when they can observe in real time how they perform relative to their own preformulated study goals and plans? I operationalize this research question by evaluating three hypotheses. These hypotheses relate to student effort, meeting study goals, and learning outcomes.

The first and main hypothesis is that students who observe their own pacesetter will exert higher study effort than students who do not observe their own pacesetter. The pacesetter is a salient moving reference point that visualizes a student's preferred study pace to reach a prespecified study goal. Falling behind the pacesetter immediately makes salient to the student that he or she is reneging on earlier commitments. Under the

[2]Clark et al. (2017) show formally that this leads to increased study effort.

assumption of reference-dependent preferences, this will be psychologically painful to the student because of loss aversion, and hence will induce the student to catch up and exert more effort. This real-time feedback mechanism is not available to students without the pacesetter. All else the same, I thus expect these control students to exert lower effort. As will be discussed in more detail in the next sections, I evaluate this hypothesis using data from a new online learning platform (MicroApp) and in-class survey responses on time investments in various learning inputs.

If students are indeed triggered not to fall behind their own pacesetter, the implication is that they will be more successful in implementing their study plan and reaching their initial study goal. The second hypothesis is therefore that students who observe their own pacesetter are more successful in reaching their study goals than students who do not observe their own pacesetter. That is, the former group has higher goal completion rates. I will evaluate the second hypothesis using the MicroApp data and detailed weekly data on study goals and plans, elicited through the in-class surveys.

Arguably, the ultimate aim of increasing student effort in a course is to have students better prepared for the exam. If the pacesetter is successful in increasing effort among students (first hypothesis), then I would expect this to have a positive impact on learning outcomes as well. The third hypothesis therefore reads as follows: students who observe their own pacesetter will exert higher study effort and as a result perform better in course examinations than students who do not observe their own pacesetter. I will match the data on student effort with administrative databases to evaluate the third hypothesis.

4 Experimental design

In order to evaluate the three hypotheses in the previous section and to answer the research question, I will implement a field experiment in a large course at a Dutch university. In this section, I elaborate on the design of the experiment. I start by discussing the institutional setting and the phase-in of the MicroApp. I then discuss the experimental conditions and the randomization of students into these conditions. Finally, I cover the implementation plan and timeline, as well as the data collection procedures.

4.1 Institutional setting

The experiment is implemented during the 2018-2019 academic year in a large first-year introductory course in microeconomics at the University of Groningen, the Netherlands. The course runs during the first block of each academic year. The first block consists of an introduction week, seven lecture weeks, and an exam week. In total, 573 students are enrolled in the course. The majority of these students are regular first-year students (about 76%). The other students are either minor finance students (12%) or pre-MSc students (12%), for whom the course is a mandatory elective.[3]

The course has the following setup. Each lecture week, students participate in a lecture, tutorial, and practical. The lecture is on Tuesday and covers the main course topics of that week. During tutorials on Wednesday or Thursday, a lecturer discusses selected exercises from the textbook. Practicals are on Friday and are led by teaching assistants. Each practical starts with a short multiple-choice quiz on the material of that week. During the second part of the practical, students work on practice exercises.

About halfway through the course, there is a midterm exam that is worth 30% of the final grade. In the exam week, there is a final exam that is worth 60% of the final grade. The remaining 10% of the final grade is the average of the six best practical quizzes (out of seven quizzes in total). In order to receive the remaining 10%, a student needs to participate in at least five (six) out of seven tutorials (practicals).

4.2 Phase-in of the MicroApp

The experiment makes use of a novel online learning platform, called the *MicroApp*, that is newly introduced in the course. The MicroApp (plus content) is developed by myself with the aim of offering students an online platform where they can practice and review the course material at their own pace and level. Students have their own password-protected account. They can design and implement individual self-study plans by selecting the

[3]Minor finance students are senior students who come from an educational program that does not give direct access to the MSc Finance program at the Faculty of Economics & Business (FEB), University of Groningen. These students first need to complete a minor program successfully before they can be enrolled in the MSc Finance program. The minor consists of courses, including the microeconomics course discussed in the text, that will help students overcome the deficiencies. Pre-MSc students are senior students who come from a different faculty or institution. These students first need to successfully complete a tailored Pre-MSc program before they can enter the MSc program of their choice at the FEB. The microeconomics course is often a mandatory component of these Pre-MSc programs.

course topics they want to work on. The MicroApp has an algorithm that automatically matches these selections with the available practice and recap material. At the time of the study, the MicroApp had 420 practice tasks (60 tasks per lecture week) and 26 recaps that summarize the course topics. Appendix section A.1 provides a few screenshots of the pre-experimental version of the MicroApp.[4]

Following the typology in Escueta et al. (forthcoming), the MicroApp combines elements of online courses, computer-assisted learning, and support for students in dealing with behaviors that negatively impact academic performance. For the purpose of this study, behavioral support will be limited to the interventions of the experiment, which are discussed in Section 4.3. An advantage that comes with the inhouse development of the MicroApp is that the interventions can be administered and precisely evaluated through A/B-testing.

The MicroApp will be offered free-of-charge to the students and is not a graded component in the course. This is because I would like to have a measure of student effort as a pure learning input, and not a measure of actions undertaken to directly satisfy some course criterium. Usage of the MicroApp is recorded in real time and saved in individual user logs, allowing me to pin down for every student when, how long, and in what way they made use of the platform. Student identifiers enable matching with administrative databases on background characteristics and test scores.

Students are notified during the introductory lecture of the course that, since the MicroApp is a new and work-in-progress educational technology, different versions may be made available and tested during the course in order to find the "optimal user experience". It is made clear that everyone will have access to the same course content on the MicroApp and that there are only slight variations in the lay-out and additional features. Students are notified that there will be weekly in-class surveys on time-use during the practicals for "general research purposes" and for improving the course. Finally, it is communicated that it is completely determined by chance which MicroApp version and in-class survey is made available to each student.

What is *not* communicated to the students during the introductory lecture is that the experimental variation is only introduced *after* the midterm exam. That is, all students

[4]This version was pre-tested in a similar microeconomics course at another faculty of the University of Groningen. Students gave the platform a perfect score (4 out of 4) in the course evaluation. This indicates that students see the added value of the platform.

are initially granted access to the version shown in Appendix section A.1. There are two reasons for a delayed start of the experiment. First, students need some time to become acquainted with the course structure and material (including the MicroApp). Since the microeconomics course is a first-year course offered at the start of the academic year, this reason holds particularly true in the setting of this study. Second, a delayed start allows me to assess MicroApp usage in the absence of interventions. For example, is there bunching of student effort right before the midterm exam (a possible indication of self-control problems)? This first stage of the study yields a convenient baseline measure of student effort.

The in-class survey administered during the practicals before the midterm exam is included in Appendix section A.2. The survey has two questions. The first question is about the microeconomics course in which the MicroApp is implemented. It deals with a student's time investment in various learning inputs, such as attending classes, usage of the MicroApp, reading offline course material, et cetera. The second question concerns a student's time investment in the other courses in which he or she participates. Responses to these questions allow me to examine the relative magnitude of each input and whether substitution between the inputs takes place. Since the survey is administered every week, I can see how time investments evolve throughout the course.

4.3 Experimental conditions

In the second stage of the study, I use the MicroApp and the in-class surveys to implement a field experiment in which I explore how students can be best motivated to stick to their initial study plan.[5] I consider three experimental conditions, which are summarized in Table 1. In all three conditions, a *broad goal* regarding MicroApp usage in the upcoming week is elicited through an additional section in the in-class survey on time-use. This broad goal indicates for each student the ex-ante preferred study time on the platform in the upcoming week. In the second and third condition, I additionally ask students in the survey to formulate a detailed *study plan*. In this plan students indicate how they are going to allocate the hours over the week in order to reach the study goal. In the third condition, the study plan is used as input to construct a real-time virtual *pacesetter* on

[5] Using the typology in Harrison and List (2004), the experiment in this study is best described as a natural field experiment.

the MicroApp. The pacesetter ("rabbit") is an individualized moving reference point that moves exactly according to the study plan submitted by a student. Appendix section A.3 shows the experimental variation in the in-class surveys. The relevant survey responses are processed the same day in the MicroApp and are made available to students the next day on.

Table 1: Summary of Experimental Conditions

	Features		
	Broad goal	Study plan	Pacesetter
Condition I: control condition	✓		
Condition II: "planner" condition	✓	✓	
Condition III: "rabbit" condition	✓	✓	✓

Notes: *Broad goal* refers to whether students are asked in weekly in-class surveys on time-use to formulate a goal with regard to how many hours they intend to use a new online learning platform (the MicroApp) in the upcoming week. Some students are also asked to formulate a *study plan* in which they make clear how they are going to allocate the hours over the week in order to reach the study goal. The *pacesetter* feature is an individualized moving reference point that moves exactly according to the study plan of the student.

The benefit of this build-on experimental design is that I can examine the marginal impact of formulating a study plan and observing the pacesetter beyond the effect of simply setting a broad study goal for the week (as in the control condition). This greatly facilitates the understanding of what will drive the results. For example, a comparison between the outcomes in the control condition and the planner condition makes it possible to pin down the stand-alone effect of formulating a detailed study plan for the week. In a similar way, I can determine the stand-alone effect of the pacesetter.

In the first experimental condition, the *control* condition, students only have access to a baseline version of the MicroApp. The baseline version contains all recaps and practice material (including feedback), but only general information on progress toward the stated study goal. Students in this condition complete a short version of the in-class survey on time-use, which does not elaborate on the study plan for the upcoming week. Figure 1 shows how the control condition is visualized in the MicroApp.

In the second condition, the *planner* condition, students do receive a request in the in-class survey to write down their study plan for the upcoming week with regard to MicroApp usage. Specifically, I ask the students to formulate daily study goals in hours and how they wish to allocate these hours over the day. The planner version of the MicroApp is the same as the baseline version, but additionally contains information on the study plan. Figure 2 shows the planner version of the MicroApp.

The third condition, the *rabbit* condition, is the same as the planner condition with the exception that the study plan is used to construct an individualized virtual pacesetter. The rabbit version of the MicroApp visualizes the pacesetter in real time as a feedback feature. Figure 3 shows the rabbit version of the MicroApp.

4.4 Randomization and exclusion rules

All students who are enrolled in the course (573 students in total) are randomly allocated to one of the three experimental conditions. The randomization is stratified based on the study phase of the students. As mentioned in Section 4.1, most students are regular first-year students, but some are senior students enrolled in a Pre-MSc program or minor finance program. Panel A in Table 2 shows that the randomization is successful in balancing the study phase composition in the three experimental conditions.

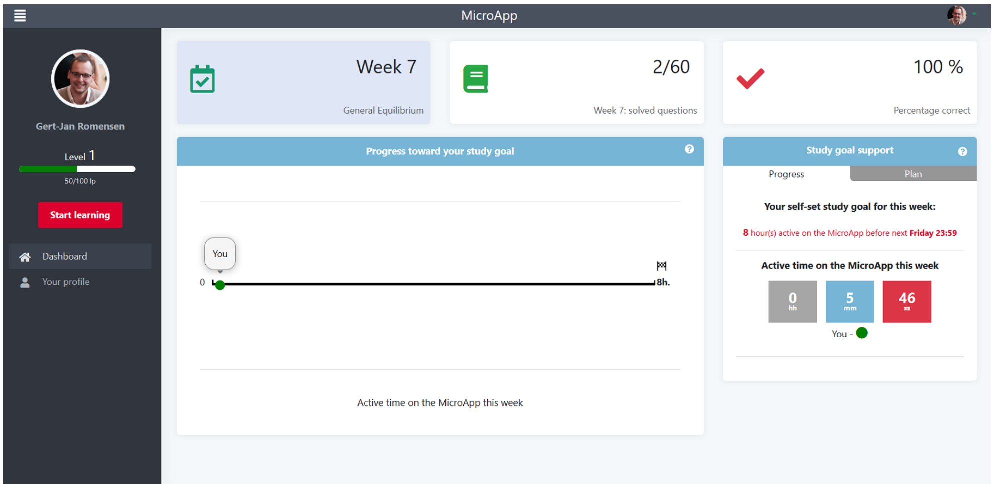
Table 2: Study Phase Composition of Students in the Exp. Conditions

		Study phase		
	All students	Regular (first-year)	Minor finance	Pre-MSc
	Panel A: sample based on all enrolled students in the course			
Condition I: control condition	190	145	22	23
Condition II: "planner" condition	191	145	22	24
Condition III: "rabbit" condition	192	145	23	24
Number of students	**573**	**435**	**67**	**71**
	Panel B: sample after imposing exclusion rules			
Condition I: control condition	104	73	15	16
Condition II: "planner" condition	99	68	14	17
Condition III: "rabbit" condition	106	75	14	17
Number of students	**309**	**216**	**43**	**50**

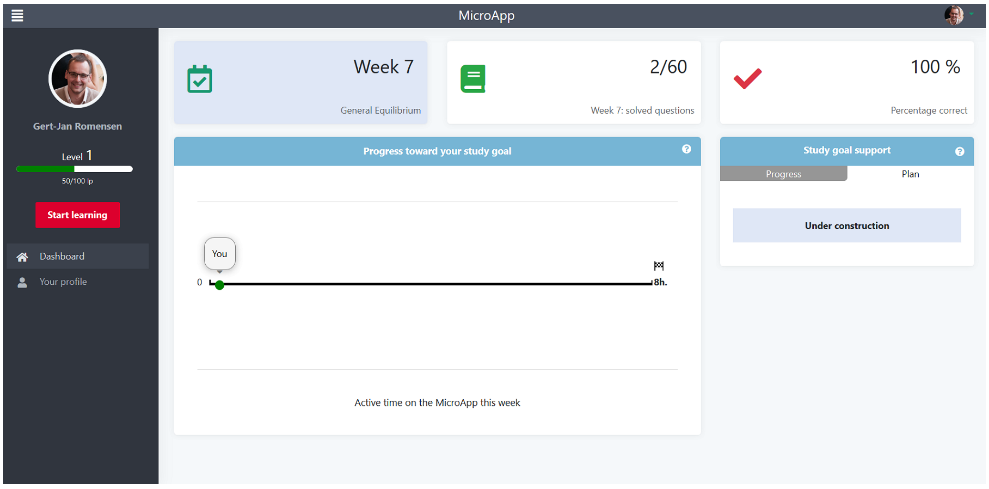
Notes: All students enrolled in a first-year introductory course in microeconomics at the University of Groningen are randomly allocated to one of the three experimental conditions discussed in Section 4.3. The randomization is stratified along the study phase of the enrolled students. Most students are regular first-year students. Some students are senior students who take the course as a mandatory elective in the minor finance program or Pre-MSc program. Please refer to footnote 3 in this chapter for further information on these two programs. The sample used in the empirical analyses is obtained after applying two exclusion rules. First, a student is excluded when he or she participated in less than two practicals after the midterm exam. Second, a student is also excluded when he or she never made use of the MicroApp in the period until the first practical after the midterm exam.

I exclude a student from the empirical analyses when he or she meets one of the following two criteria. These criteria were specified in the pre-analysis plan of this study. First, the

Figure 1: MicroApp Version for Students in the Control Condition



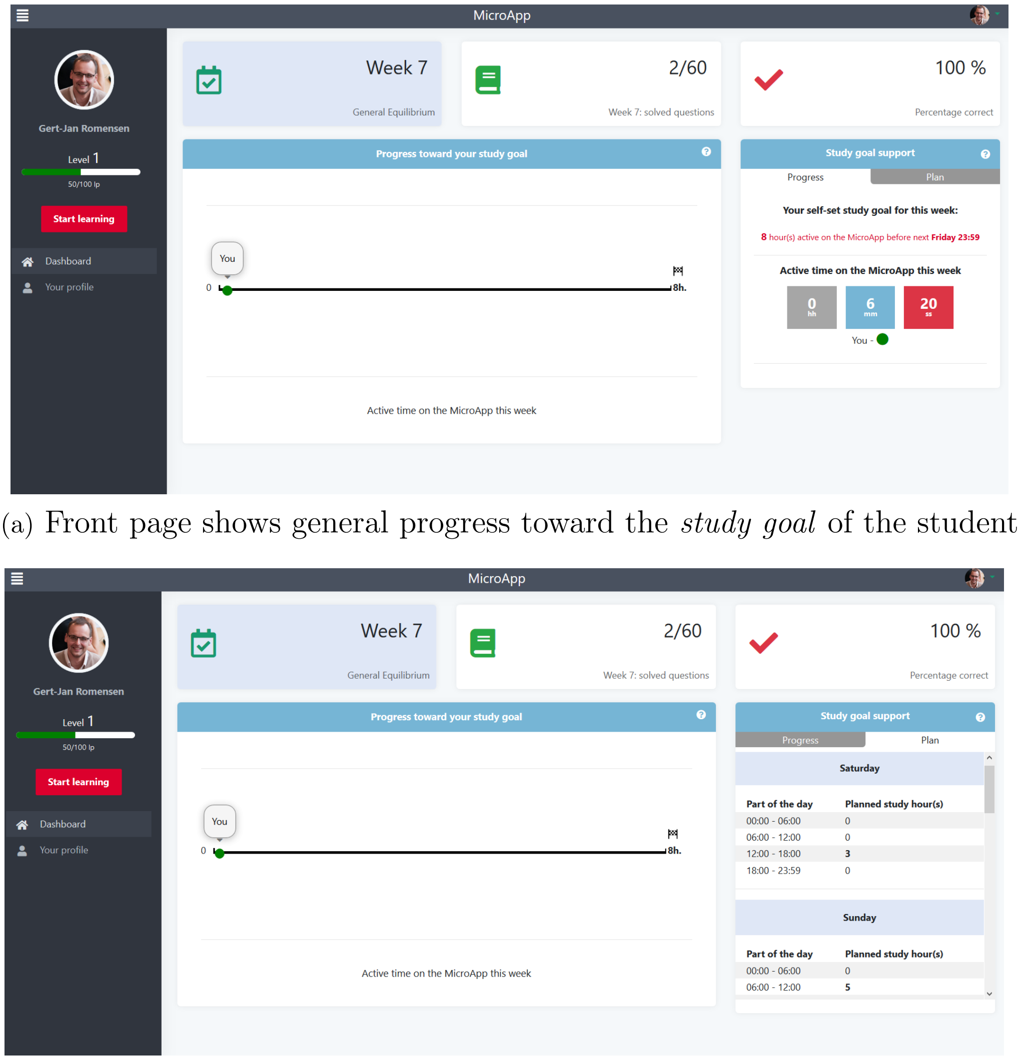
(a) Front page shows general progress toward the *study goal* of the student



(b) The *study plan* feature is not available for students in the control condition

Notes: Students in the control condition have access to a baseline version of the MicroApp (a new online learning platform in the course). The baseline version only shows general progress toward the study goal of the student. The study goal is elicited weekly in in-class surveys and is about intended MicroApp usage in the upcoming week. In the control condition, students are not requested to submit a study plan in which they indicate how they are going to allocate study hours over the week in order to reach the study goal. The study plan feature is therefore not available in the MicroApp ("under construction").

Figure 2: MicroApp Version for Students in the Planner Condition

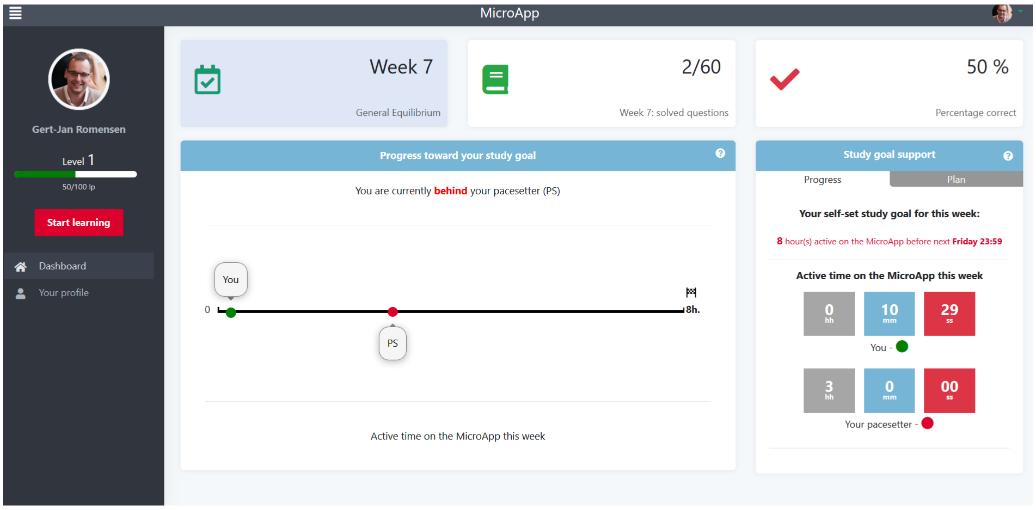


(a) Front page shows general progress toward the *study goal* of the student

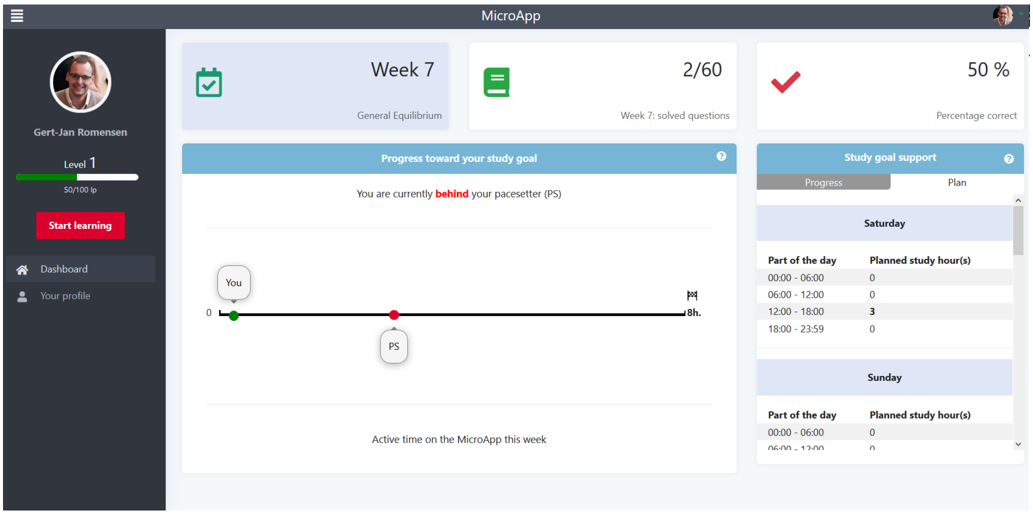
(b) The *study plan* feature is available for students in the planner condition

Notes: Students in the planner condition have access to the same version of the MicroApp as students in the control condition. The only difference in the MicroApp being that students in the planner condition can request and review their initially submitted study plan at any time. This study plan is elicited weekly in the in-class surveys, together with the corresponding study goal. The study goal is about intended MicroApp usage in the upcoming week and the study plan is about how study hours are allocated over the week in order to reach the study goal.

Figure 3: MicroApp Version for Students in the Rabbit Condition



(a) Front page shows general progress toward the *study goal* of the student as well as progress relative to a student's own pacesetter (*rabbit*)



(b) The *study plan* feature is available for students in the rabbit condition

Notes: Students in the rabbit condition complete the same in-class survey questions and have access to the same version of the MicroApp as students in the planner condition. The only difference in the MicroApp is that in the rabbit condition students additionally can track their study progress relative to their own pacesetter (rabbit). The pacesetter is an individualized moving reference point that moves exactly according to the study plan submitted by a student.

student participated in less than two practicals after the midterm exam. The reason for this first criterium is because the interventions are based on input from the in-class surveys administered after the midterm exam. When a student is frequently absent during these classes, the interventions cannot be based on up-to-date information from the students. Second, the student never made use of the MicroApp in the period until the first practical after the midterm exam (when the first in-class survey with experimental variation was administered). With no previous user experience, a student cannot formulate a well-grounded study goal and plan with regard to MicroApp usage. I consider a student having used the MicroApp when he or she solved at least one task on the platform.

Panel B in Table 2 shows that balance in terms of study phase composition is preserved when the two exclusion rules are applied. About 83% of the students participated in at least two practicals after the midterm exam, of which 65% were also active users of the MicroApp. All data and results in the remainder of this chapter are based on this restricted sample (unless stated otherwise), as these are the students who actively participate in the course and also make use of the MicroApp. As will be discussed in Section 5.1, additional balance tests reveal no significant differences between the experimental groups in terms of a wide range of pre-experimental characteristics, indicating that the randomization is successful.

4.5 Implementation plan and timeline of the experiment

Table 3 summarizes the timeline of the experiment by listing all notable events. The experiment consists of two stages. In the first stage, covering the period before the first practical class after the midterm exam, there is no experimental variation in the course. All students have access to the same version of the MicroApp and complete the same in-class surveys (which do not yet elaborate on weekly study goals and plans).[6] Friday 5 October 2018 marks the transition between the first stage and the second stage. On this day with practical classes, the in-class surveys feature experimental variation for the first time. The second stage begins on Saturday 6 October 2018 (start of lecture week 5), as experimental variation is now introduced in the MicroApp as well. That is, in the second stage the experiment is in full swing, with treatment-specific versions of the surveys and

[6]The pre-experimental version of the MicroApp is shown in Appendix section A.1. The standard format of the in-class survey used in the first stage is shown in Appendix section A.2.

the platform.

The MicroApp was introduced to the students during the introductory lecture of the course, along with a few general comments that were discussed in Section 4.2. Personal login credentials were shared with students the next day via email. The actual recording of the time spent on the MicroApp (my measure of student effort) started a few days later in order to let it coincide with the start date of the regular lecture weeks in the course. I decided to do so because the introduction week (the week prior to the regular lecture weeks) is a non-standard week that is mainly reserved for information meetings, workshops and social events. These activities leave students with little time to engage in actual study activities.

Table 3: Timeline of the Experiment

Date	Event(s)
3 September 2018 - 4 October 2018	**First stage (no experimental variation)**
3 September 2018 - 7 September 2018	Introduction week at the university (start of the academic year)
4 September 2018	MicroApp introduced during an introductory lecture
5 September 2018	Personal MicroApp login credentials sent to students via email
10 September 2018	Start of regular lecture weeks in the course
	Start of recording the time spent on the MicroApp
14 September 2018	First in-class survey on time-use
	Extra questions in first survey on overall goals in the course
21 September 2018	Second in-class survey on time-use
28 September 2018	Third in-class survey on time-use
	Third survey includes a task with intertemporal choices about real effort
2 October 2018	*Midterm exam*
5 October 2018	Transition day. Fourth in-class survey on time-use
	The fourth and remaining surveys come in three versions
	Each experimental condition completes its own version
6 October 2018 - 1 November 2018	**Second stage (experimental variation)**
6 October 2018	From now on, three different MicroApp versions online
	Each experimental condition has access to its own version
12 October 2018	Fifth in-class survey on time-use
19 October 2018	Sixth in-class survey on time-use
26 October 2018	Seventh in-class survey on time-use
1 November 2018	*Final exam*

The surveys on time-use were completed in-class during the practicals on Friday. To ensure unique identification and a high response rate in the first stage, I attached the surveys to the (graded) weekly practical quizzes. Students first completed the quiz and afterwards were allocated extra time to complete the survey. Students work on the quizzes and surveys individually (no talking allowed) and write down their name and/or student number once both are completed.[7]

[7]The filled-in surveys were collected by teaching assistants and afterwards handed over to me for

In addition to the two core questions on time investments in learning inputs, the first-stage surveys were used to obtain some further information on the initial motivation of the students and the presence of self-control problems. In the first survey, I asked the students to write down their goals in terms of the overall course grade, grade on the midterm exam, and grade on the final exam. I also asked each student to indicate the chances of him or her participating in the final exam.[8] The precise wording of the questions can be found in Appendix section A.4. In the spirit of Augenblick, Niederle and Sprenger (2015), the third survey was in part used to administer a task that aims to measure whether students exhibit present biased preferences in real-effort decisions. The task closely follows the version used in Carvalho, Meier and Wang (2016). In the task, students have to indicate whether they wish to complete a long online research survey at a later date or a shorter version of the same survey at an earlier date. The question is asked repeatedly with variations in two dimensions: the length and the deadline of the shorter survey. The precise wording of the task can be found in Appendix section A.5. The responses to these additional questions and tasks are discussed in Section 5.1.

Students were notified beforehand during the first lecture after the midterm exam that "some new features" are being tested during the second part of the course. They were made aware that this testing implies that different versions of the MicroApp and the in-class surveys are going to be used.[9] Students were ensured that everyone would continue to have access to the same course material on the MicroApp. It was also made explicitly clear to each student that whichever version is made available is completely determined by chance. Teaching assistants were trained on how to respond in case students had questions

further processing.

[8]Asking a student to indicate the chances of him or her participating in the final exam serves as an indication for how committed the student is to complete the course. In the Netherlands, students sometimes take a course as a non-mandatory elective and in such cases the commitment to finish the course may not always be high.

[9]I slightly changed the procedure during the practicals in order to accommodate the experimental variation in the in-class surveys. I detached the surveys from the quizzes and instead placed them in three separate plastic folders. Each folder contained one version of the in-class survey as well as a list with students to whom the survey should be handed out. The lists follow from the randomization procedure in Section 4.4. All students (panel A in Table 2) are included in the lists. The folders were shared with the student assistants prior to the start of each practical class. The class started with the quiz, followed by a short break. Afterwards, teaching assistants distributed the surveys by going through the lists and calling the student names. Students completed the surveys individually (no talking) and wrote down their name and/or student number on the survey. The latter made it possible to uniquely match survey responses with MicroApp accounts and administrative databases.

about the new procedure.[10] This general information was provided to the students in order to avoid that the new procedure and the new versions come as a complete surprise.

The post-midterm survey responses are processed the same day in the MicroApp. For each student, I enter the MicroApp study goal for the upcoming week (all experimental conditions) and the study plan for the upcoming week (conditions II and III). The MicroApp automatically updates all accounts and versions with the new information. From the next day on, the information is displayed in the platform and students can at any time track their progress relative to their own goal and plan.[11] This procedure is repeated every lecture week until the final exam is administered. In case a student did not submit survey responses during a particular week, the information segments in the platform will remain empty.

4.6 Data collection

I operationalize study effort by focusing on the time spent on the MicroApp. The interventions of the experiment are geared toward helping students implementing their initial study plan with respect to this effort measure. Of course, I recognize that overall student effort is not limited to time spent on the MicroApp and may also involve time investments in other learning inputs, such as attending classes and reading the textbook. The issue with these other components of student effort, however, is that they are either difficult to measure (e.g., reading the textbook) or a mandatory aspect of the course (e.g., "participation grade" for class attendance). This makes these other measures of student effort less suitable for this study.

There are three reasons for focusing on the time spent on the MicroApp rather than, say, the number of solved practice tasks on the platform. First, time spent is a measure less confounded by ability. Second, effort on the MicroApp may also, among other things, entail reading feedback and/or recaps. This would not be well-captured by the alternative measure. Third, by focusing the study goal on the time spent on the platform, rather than the number of solved tasks, I avoid creating perverse incentives for skipping feedback and quickly solving many tasks by submitting random answers.

To verify whether the time spent on the MicroApp is used productively, fifteen minutes

[10] In case of very specific questions or an unwillingness to further participate in the surveys, a student could at any time contact the secretary of the course. No student made use of this option.

[11] In Section 4.3, I discussed how the information is displayed in the MicroApp.

of inactivity (no mouse movements) will activate a pop-up requesting the student to respond by clicking on a button. In case of no response to this pop-up, the student will be automatically logged out after one minute and the counting of time will stop for that particular session. This procedure ensures that the measure of student effort captures active engagement with the course content on the platform. It also rules out that a student can meet his or her study goal by passively staying logged in.

To assess the relative magnitude of other learning inputs in students' study efforts, the in-class surveys elicit weekly time investments in the following seven categories: (1) attending classes; (2) reading offline course material (e.g., textbook, lecture slides, handouts); (3) offline practice (e.g., tutorial exercises and old exams); (4) use of the MicroApp; (5) reading/watching online course material (not on the MicroApp); (6) online practice (not on the MicroApp); (7) any other study activity that is not captured by the previous categories (e.g., attending office hours). Students can tick per category one of the available answer boxes, varying from "not at all" to "five hours or more than five hours".

To complete the dataset used in this study, I match the MicroApp data and the in-class survey responses with administrative databases that contain further student background information. Specifically, I obtain student-level data on gender, age, average grade in prior education, and the type of prior education. Through the administrative databases, I also collect information on class attendance and interim course performances.

5 Empirical evaluation

This section presents the empirical evaluation and discusses the findings. I start by performing balance tests on the first-stage data (baseline). The second stage (experimental period) is discussed afterwards in four different parts. In the first part, I examine how the interventions affect student effort (hypothesis I). I will distinguish between effort on the MicroApp and time investments in other learning inputs. The second part discusses goal completion rates (hypothesis II). Third, I cover the treatment effects on learning outcomes (hypothesis III). In the final part, I perform a more exploratory analysis that looks at the effects of learning inputs.

Throughout the discussion of the second stage, I will present the overall treatment effects as well as the treatment effects for the following three subgroups: males/females, high-performers/low-performers, and procrastinators/nonprocrastinators. These groups were prespecified in the pre-analysis plan of this study. I operationalize high/low pre-experimental performance as the within-treatment relative performance of a student on the midterm exam. Low-performance (high-performance) students are in the within-treatment bottom (top) 50% in terms of midterm performance.[12] Procrastinators are identified based on course enrollment dates (as will be discussed later).

5.1 First stage: no experimental variation

During the first stage of the study, no experimental variation is introduced. All students have access to the same version of the MicroApp and complete the same in-class surveys. Table 4 provides the first-stage descriptive statistics. Note that, as in the remainder of this chapter (unless stated otherwise), the table reports on the students that satisfy the inclusion rules of this study (covered in Section 4.4). The responses from the fourth in-class survey (administered on 5 October 2018 (transition day)) are also included in this table. This is because the survey questions relate to the fourth lecture week (a first-stage week).

Students are on average about 20 years old at the start of the course. Most are in the first year of their study (70%). The share of female students in the sample is 30%. The majority of the students, about 75%, completed their pre-university education in the Netherlands. Restricting attention to the Dutch regular first-year students, as these share the same pre-university educational background, the average pre-university grade is 6.82 out of 10. Demographics appear balanced across the three experimental conditions.

Students are in general committed to completing the course and rate the chances of participating in the final exam at roughly 97%. The students aim for an overall course grade of 7.69 (out of 10). Performance goals are similar for the midterm exam (7.94) and the final exam (7.61). The initial course goals of the students are similar across the three conditions.

Focusing on pre-experimental behavior on the MicroApp, I again observe that students

[12] In the pre-analysis plan (PAP), I operationalized initial ability as the average pre-university grade of a student. After completing the PAP, however, I was informed that this information is only available for the Dutch students in the sample. Using this operationalization would thus exclude all international students from the analysis (about 25% of all enrolled students). Furthermore, heterogeneity in the type of pre-university education hinders fair comparisons of pre-university grades. For these reasons, I will use the alternative operationalization discussed in the main text.

Table 4: First-Stage Descriptive Statistics of Experimental Conditions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Full sample		**T1: Control**		**T2: Planner**		**T3: Rabbit**		Joint test: treatment effect = 0	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.		
	Pre-experimental behavior on the MicroApp									
Avg. time on the app (in hours)	6.14	(6.43)	6.04	(6.57)	6.07	(6.42)	6.30	(6.36)	0.05	[0.95]
Number of solved tasks	69.49	(61.20)	70.69	(63.59)	70.48	(61.38)	67.37	(59.12)	0.10	[0.91]
Time needed to solve a task (in minutes)	2.76	(1.73)	2.62	(1.59)	2.87	(1.91)	2.79	(1.68)	0.58	[0.56]
Share of tasks solved correctly (%)	61.74		61.36		62.44		61.44		0.14	[0.87]
Recap activities	18.40	(16.74)	18.42	(16.60)	17.35	(15.94)	19.38	(17.71)	0.34	[0.71]
	Pre-experimental behavior in the course									
Grade on practical quizzes (out of 10)	7.50	(1.04)	7.47	(1.03)	7.50	(0.98)	7.53	(1.11)	0.07	[0.93]
Score on the midterm exam (out of 30)	21.02	(5.21)	21.28	(5.32)	20.92	(5.54)	20.86	(4.80)	0.20	[0.82]
Number of tutorials and practicals attended	7.43	(0.84)	7.45	(0.81)	7.37	(0.90)	7.46	(0.81)	0.31	[0.73]
	Demographics									
Age at the start of the course (in years)	20.49	(2.23)	20.30	(2.17)	20.65	(2.12)	20.53	(2.40)	0.66	[0.52]
Avg. pre-university grade (out of 10)	6.82	(0.53)	6.80	(0.55)	6.86	(0.55)	6.80	(0.50)	0.23	[0.80]
Share of female students (%)	30.26		33.66		30.61		26.67		0.60	[0.55]
	Study phase (# students)									
Regular first-year students	216		73		68		75		0.05	[0.95]
Minor finance students	43		15		14		14		0.06	[0.94]
Pre-MSc students	50		16		17		17		0.04	[0.96]
	Initial course goals of students									
Overall course grade (out of 10)	7.69	(0.88)	7.77	(0.94)	7.68	(0.91)	7.62	(0.80)	0.56	[0.57]
Midterm exam (out of 10)	7.94	(0.95)	8.02	(0.90)	7.86	(1.04)	7.95	(0.89)	0.53	[0.59]
Final exam (out of 10)	7.61	(1.03)	7.76	(1.09)	7.59	(0.97)	7.51	(1.01)	1.13	[0.32]
Chances of taking the final exam (out of 100%)	96.64		97.93		96.90		95.32		1.50	[0.22]
	Pre-experimental median responses to in-class surveys									
Time spent on course learning inputs (per week)										
Attending classes	5		5		5		4		0.81	[0.45]
Reading offline material	3		3		3		3		0.33	[0.72]
Practicing with offline material	3		3		3		2		5.01	[0.01]
Use of the MicroApp	2		2		2		2		0.15	[0.86]
Reading/watching online material (not MicroApp)	1		1		1		1		1.58	[0.21]
Practicing with online material (not MicroApp)	1		1		1		1		0.57	[0.57]
Other (not captured by the above)	1		1		1		1		5.57	[0.00]
Avg. time spent on other courses (per week)										
Other courses	3		3		3		3		2.31	[0.10]
Number of students	309		104		99		106			

Notes: Unit of observation is the student. Columns (9) and (10) show F-statistics and [p-values] from a balance test of whether the treatment coefficients are jointly equal to zero. Students' demographics and study phase are retrieved from administration records of the university. The average pre-university grade is based on data from Dutch regular first-year students, as these students have the same pre-university educational background. Minor finance students and Pre-MSc students are senior students who follow a tailored deficiency program (of which the microeconomics course is a mandatory component) before they are allowed to enroll in a MSc program at the university (see footnote 3 in this chapter for further details). Initial course goals of the students are collected during the first in-class survey in the course. With regard to the survey question related to the weekly time investments in course learning inputs, students could tick one of the following available answer boxes: (1) not at all, (2) less than 1 hour, (3) 1− < 3 hour(s), (4) 3− <5 hours, (5) 5 hours or more than 5 hours. The question on the average weekly time investment in other courses had the following answer boxes: (1) not at all, (2) less than 5 hours, (3) 5− < 10 hour(s), (4) 10− <15 hours, (5) 15 hours or more than 15 hours.

in the different conditions behave in the same way. Students on average spent six hours studying on the MicroApp during the first stage. Each student solved about 70 practice tasks, of which 62% was answered correctly. A little less than three minutes was spent on each task, which suggests that students at least took the time to read the question and did not simply click through in an inattentive manner. Students loaded recap modules (clicked on the start button) about eighteen times on average, which is a rough indication of having read the respective recaps.

Performances in other aspects of the course display the same pattern. Students scored an average of 21 points (out of 30) on the midterm exam, which translates in a grade of 7 (out of 10). Similar results are found for the weekly practical quizzes, on which the students scored a 7.5 (out of 10). Class attendance was high during the first stage, which indicates high involvement in the course. Although this may in part be due to attendance being incentivized through a participation grade.

Considering the pre-experimental responses to the in-class surveys, we get an indication of how students spent their time in the first part of the course. Per learning input students could tick one of the five available answer boxes with weekly time investments, varying from *not at all* to *5 hours or more than 5 hours.*[13] Table 4 shows that across all conditions the most popular learning inputs are attending classes, reading offline, practicing offline, and use of the MicroApp.

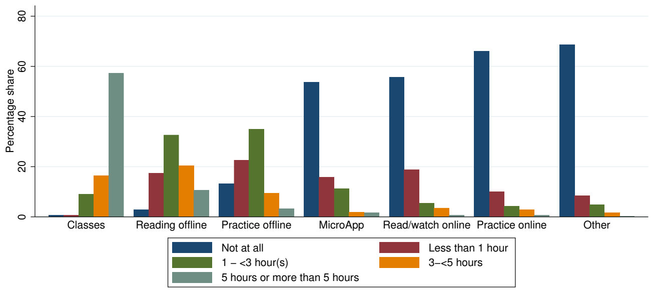
Figure 4 aggregates per lecture week the first-stage time investments in learning inputs. This figure reveals that time investments change during the first-stage lecture weeks. Early on in the course, students spent most of their time "consuming" course material (e.g., attending classes and reading), whereas later on the focus shifts toward practicing. Practicing is mainly done with offline material (e.g., old exams) or online on the MicroApp. Few students indicate that they make use of other online resources. The surveys also show that the average time investment in other courses is similar across the three groups (5-10 hours per course).[14]

Overall, the previous discussion indicates that students in all three experimental conditions are similar in terms of observable characteristics in the pre-experimental period.

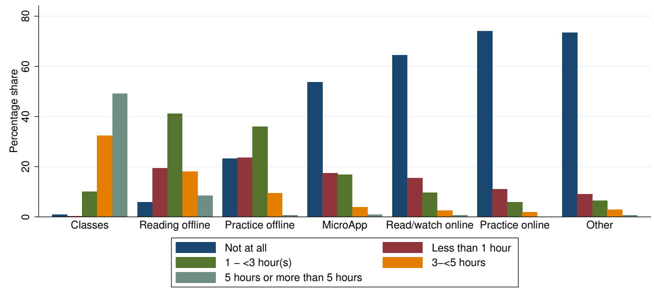
[13]The full choice set was as follows: (1) not at all, (2) less than 1 hour, (3) $1-<3$ hour(s), (4) $3-<5$ hours, (5) 5 hours or more than 5 hours.

[14]For the survey question related to the average time investment in other courses, the available answer boxes were: (1) not at all, (2) less than 5 hours, (3) $5-<10$ hour(s), (4) $10-<15$ hours, (5) 15 hours or more than 15 hours.

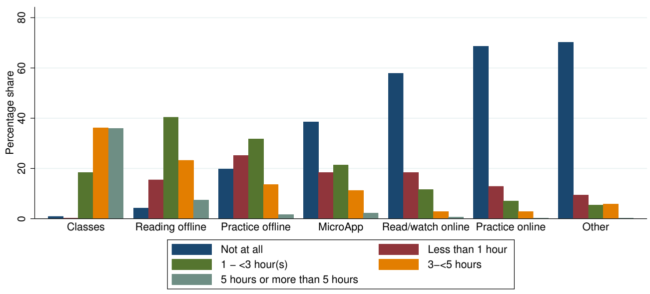
Figure 4: First-Stage Weekly Time Investments in Learning Inputs



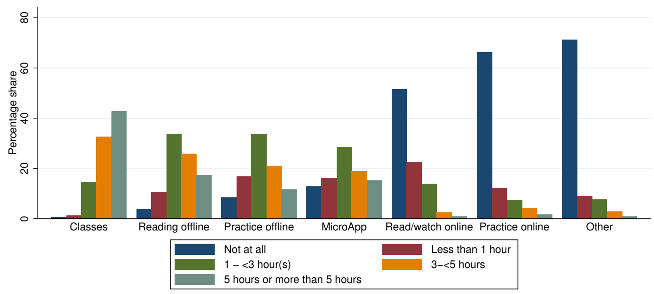
(a) Lecture week 1



(b) Lecture week 2



(c) Lecture week 3



(d) Lecture week 4

Notes: Students' time investments in learning inputs are elicited at the end of each lecture week via in-class surveys. In the first stage of the experiment (overlapping with the first four lecture weeks) there is no experimental variation, which means that all students have access to the same version of the MicroApp and the in-class surveys. Halfway through lecture week 4 students participated in a mid-term exam (worth 30% of the final grade). The learning input "classes" is partly incentivized (attendance is required to be eligible for a participation grade (worth 10% of the final grade)). The MicroApp is a newly introduced online learning platform in the course with abundant practice- and recap material. The categories "read/watch online" and "practice online" refer to non-MicroApp online study activities.

This suggests that the randomization is successful in creating balance before experimental variation is introduced, and that this result is robust to the applied exclusion rules.

Do we observe present-biased behavior among students prior to the start of the experiment? In one pre-experimental in-class survey, students completed an incentivized multiple price list with intertemporal choices about filling in a shorter online survey earlier or a longer online survey later.[15] Note that these choices relate to real effort rather than monetary rewards. This is for two reasons. First, it more closely resembles the decision problem of a student in which I am interested, namely the decision how much study effort to exert at any given time. Second, intertemporal choices about monetary rewards may not be well-suited to capture dynamic time-inconsistent preferences due to several confounds (such as payment reliability). After controlling for these confounds, recent studies find no or limited evidence of time inconsistency in monetary choices (Giné, Goldberg, Silverman and Yang 2018, Andreoni and Sprenger 2012, Andersen, Harrison, Lau and Rutströmd 2014). Augenblick et al. (2015) also report a lack of evidence of time inconsistency in monetary decisions, but they do find considerable present bias in decisions by students that are based on real effort.

In the multiple price list method, I closely follow Carvalho et al. (2016) and let students choose between soon completing a short survey or later completing a long survey, where I vary the length of the shorter survey. The choices are presented to students in two time frames: a short time frame (5 days (sooner) x 35 days (later)) and a long time frame (90 days (sooner) x 120 days (later)). There are ten choices in total, five per time frame. Similar to the approach in Meier and Sprenger (2015), I elicit individual discount rates (IDRs) by looking at the point where a student switches from the short survey to the long survey in each time frame. These switching points yield interval information about students' intertemporal preferences.[16]

[15]The multiple price list can be found in Appendix section A.5. The online survey was an actual survey from a research team at the faculty (unaffiliated with this study). In line with the information stated in the multiple price list, there was a short version of the online survey and a long version. The short version, in turn, varied in length such that it would take approximately 15 minutes, 18 minutes, 21 minutes, 24 minutes or 27 minutes to complete.

[16]As in Carvalho et al. (2016), let x be the duration in minutes of the longest sooner survey a student chooses over the later survey. The (lb, ub) (lower bound, upper bound) of the discount rate intervals are then: $(15/30, 18/30)$ for $x = 15$; $(18/30, 21/30)$ for $x = 18$; $(21/30, 24/30)$ for $x = 21$; $(24/30, 27/30)$ for $x = 24$, and $(27/30, 30/30)$ for $x = 27$. For those students who always pick the later survey, I will use the interval $(0, 15/30)$. For each time frame, I respectively calculate the lower bound and upper bound of the individual discount rates (IDRs) as follows: $IDR = \frac{1}{ub} - 1$ and $IDR = \frac{1}{lb} - 1$.

Table 5: Pre-Experimental Intertemporal Choices About Real Effort

	(1)	(2)	(3)	(4)
Short time frame=1	0.041	0.056	0.047	0.048
	(0.046)	(0.042)	(0.035)	(0.032)
Treatment=2	-0.053	-0.040	-0.030	-0.028
	(0.058)	(0.052)	(0.048)	(0.043)
Treatment=3	0.087	0.074	0.078	0.050
	(0.065)	(0.061)	(0.053)	(0.049)
Short time frame=1 × Treatment=2	0.093	0.082	0.065	0.073
	(0.067)	(0.064)	(0.051)	(0.048)
Short time frame=1 × Treatment=3	-0.074	-0.041	-0.061	0.005
	(0.067)	(0.060)	(0.055)	(0.049)
Constant	0.334***	0.259***	0.319***	0.254***
	(0.044)	(0.041)	(0.036)	(0.034)
Number of students	220	188	333	286
Number of observations	440	376	666	572
Student inclusion/exclusion rules applied?	Yes	Yes	No	No
Conditional on consistent choices?	No	Yes	No	Yes

Notes: This table reports the estimates of interval regressions where the dependent variable is an interval measure of the individual discount rate (IDR). Standard errors are clustered at the student level. Students were asked to make intertemporal choices about completing a shorter survey earlier or a longer survey later. Two time frames were presented to the students: a short time frame (5 days (sooner) x 35 days (later)) and a long time frame (90 days (sooner) x 120 days (later)). IDRs are estimated for each time frame. Only students who completed all choices in the price list are considered. Choices are consistent when a student has at most one switching point in each time frame. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

In Table 5, I present the results of interval regressions where the dependent variable is an interval measure of the individual discount rate (IDR). As explanatory variables I include indicators for the short time frame and treatment status. Including treatment status serves as an additional balance check, since IDRs were elicited before experimental variation was introduced. Present bias in real effort is observed when the individual discount rate is higher in the short time frame than in the long time frame. In column (1) we observe that the point estimate of the IDR is indeed higher in the short time frame, although it is not different from zero at conventional significance levels. As expected, treatment status has no impact on IDRs. Results are similar when I only look at students who made consistent choices (about 85% of the students). I consider choices to be consistent when there is at most one switching point per time frame (which satisfies monotonicity). There is no change in outcomes when the full sample of students is considered (no exclusion rules applied). In all cases, the point estimate of the IDR is higher in the short time frame (about 5 percentage points higher), suggesting present bias, but the estimate is never significantly different from zero.

The two time frames make it possible to identify per student whether they exhibit present bias in their intertemporal effort choices. That is, when their IDR is higher in the short time frame than in the long time frame. Let K_{ST} (K_{LT}) be the switching point in the short time frame (long time frame). I classify a student as present-biased in effort when $K_{LT} > K_{ST}$. Using this definition, about 10% of the students display present bias in their choices. This reduces to 8% when only consistent choices are considered.[17]

In the pre-analysis plan, I indicated that I am interested in performing a subgroup analysis by contrasting outcomes between present-biased students and non-present-biased students. The multiple price list method, however, yields a sample of present-biased students that is too small for conducting a meaningful analysis. The small number identified may in part be due to difficulties in task comprehension and the fact that many students had no switching point at all.[18]

An alternative way of identifying students prone to procrastination is proposed by Himmler, Jäckle and Weinschenk (2019). They use the median application date to an educational program as a cutoff for distinguishing procrastinators (applied after the median date) from nonprocrastinators (applied before), reasoning that nonprocrastinators are more likely to take care of things right away (such as the application procedure) and not wait until the last minute. The application date is treated as a proxy for procrastination tendencies.[19] Just like the multiple price list method, a benefit of this approach is that it relies on actual behavior and not on self-reported measures (which may be flawed when answered in a socially-desirable way).

I will adopt a similar approach by looking at the course enrollment dates of students. Figure A.9 shows the number of students per enrollment date.[20] The most popular day for enrolling is the day of the registration deadline, which is suggestive of procrastination. The median course enrollment date in the full sample (21 August 2018) is a little less than one week before the registration deadline (26 August 2018). About 30% of the students registered after the deadline. These students were still granted access to the course and

[17]Percentages are not much different when the full group of students is evaluated.

[18]After filling in the in-class survey, some students indicated to teaching assistants that they did not fully understand the survey part related to the multiple price list. About 32% of the students had no switching point at all. That is, in all ten choices these students always chose the same version of the online survey.

[19]Other recent papers also make use of enrollment/application behavior to gauge procrastination tendencies. Reuben, Sapienza and Zingales (2015) consider the application date to an MBA program as a procrastination proxy. De Paola and Scoppa (2015) look at the time elapsed between receiving a university acceptance letter and the date of enrolling. Brown and Previtero (2014) study financial behaviors and define a procrastinator as an individual who waits until the last day of the enrollment period before making a decision which health care plan to select.

[20]The figure shows all enrolled students in the course. No exclusion rules are applied.

were given permission to take the exams.[21]

The within-treatment median course enrollment date is used as the cutoff for identifying procrastinators and nonprocrastinators in each treatment. Table 6 summarizes the results per treatment. The random allocation of students to the experimental conditions successfully balances the median enrollment date in each group. In total, 268 students (289 students) are identified as procrastinators (nonprocrastinators).

Table 6: Identifying Procrastinators in the Course

		# Students (% share)	
	Median enrollment date ($\tilde{D}$)	Procrastinators ($> \tilde{D}$)	Nonprocrastinators ($\leq \tilde{D}$)
All students	21 August 2018	268 (48.11%)	289 (51.89%)
Condition I: control condition	21 August 2018	84 (46.65%)	100 (54.35%)
Condition II: "planner" condition	20 August 2018	92 (49.20%)	95 (50.80%)
Condition III: "rabbit" condition	21 August 2018	92 (49.46%)	94 (50.54%)

Notes: The within-treatment median course enrollment date is used as a cutoff for identifying procrastinators and nonprocrastinators in each treatment. All enrolled students are considered (no inclusion/exclusion rules are applied). For 16 students the enrollment date could not be identified.

Some basic analyses (not shown) reveal that students identified as procrastinators perform slightly worse in the course prior to the experimental variation. They score, on average, about 0.6 points lower on the midterm exam (a total of 30 points could be obtained). Himmler, Jäckle and Weinschenk (2019) also observe that procrastinators perform worse in exams. They find that procrastinating students tend to pass fewer exams.

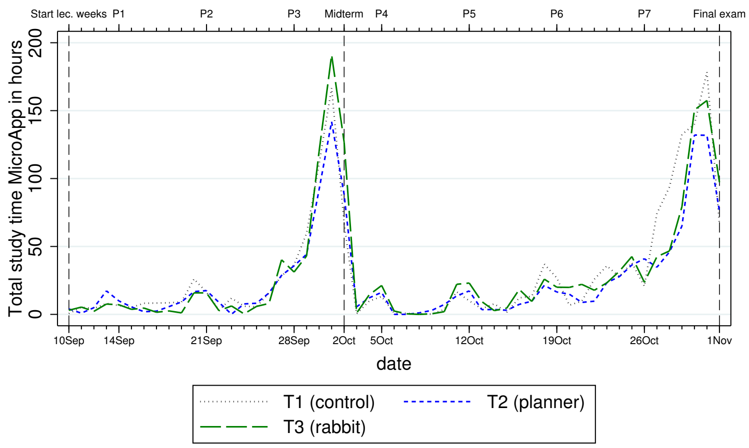
5.2 Second stage: experimental variation

Time investments in learning inputs The first hypothesis I wish to examine, as detailed in Section 3, is that students who observe their own pacesetter will exert higher study effort than students who do not observe their own pacesetter. I find support for this hypothesis when students in the rabbit condition invest significantly more time in the learning input targeted by the treatments (studying on the MicroApp). I will start by discussing study effort on the MicroApp and then continue with students' time investments in the other learning inputs.

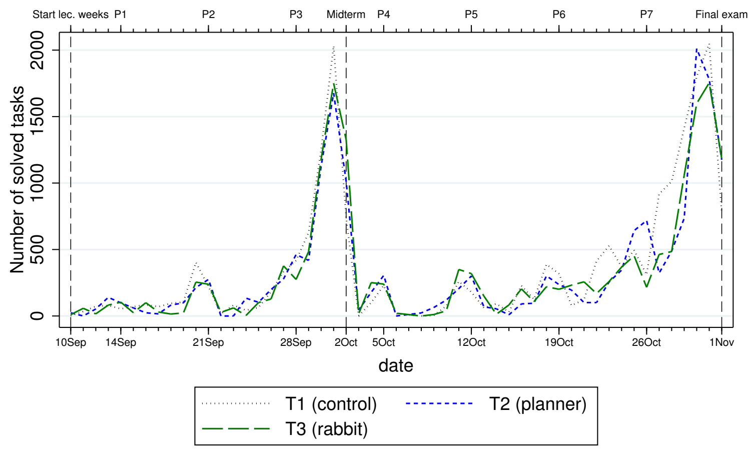
Figure 5 provides an overview of daily student effort (aggregated) on the MicroApp during the course. Panel A (panel B) of the figure displays the sum total of daily study time (solved practice tasks) per experimental condition. Both measures of student effort display the same overall pattern. In the first stage (the period before the fourth practical

[21] A further check reveals that indeed 30% of the students who took the midterm exam belong to the group of students that registered late.

Figure 5: Sum Total of Daily Study Time on the MicroApp



(a) Study time



(b) Solved practice tasks (alternative measure)

Notes: The figure shows per experimental condition the sum total of study time (panel A) and the number of solved practice tasks (panel B) on the MicroApp. The regular lecture weeks of the course started on Monday 10 September 2018. Surveys on time-use were administered during each practical class (P). Experimental variation in the surveys was introduced during the fourth practical class ($P4$) on 5 October (the first practical after the midterm exam) and continued until the final exam. Three different MicroApp versions (one per experimental condition) aired on 6 October and stayed online until the final exam. The period before (after) 5 October marks the first (second) stage of the experiment.

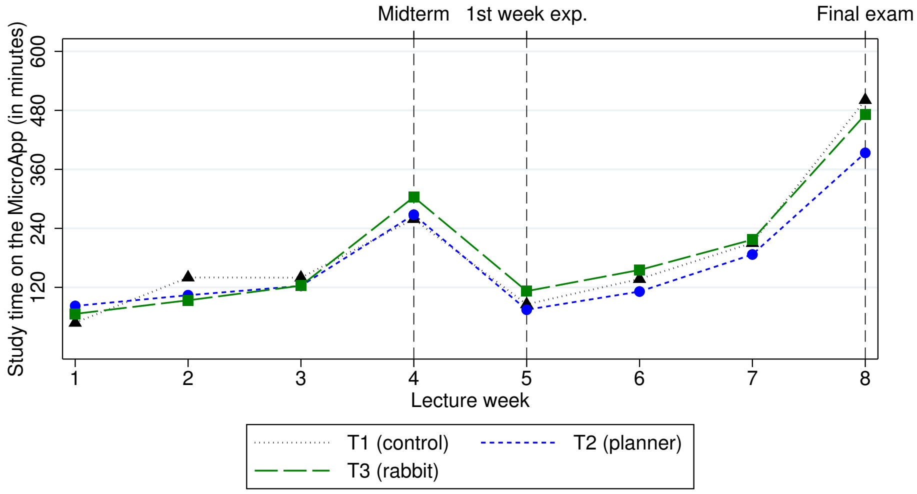
class ($P4$)), students in all three experimental conditions bunch their effort right before the due date of a graded course component, such as practical quizzes and the midterm exam. That is, overall student activity is much higher in the days right before such events. This *may* be indicative of self-control problems on behalf of the students. Students may therefore see potential in using the planning tools provided in the interventions to get starter earlier on preparing for the graded tasks. Consistent with the discussion in Section 5.1, I observe no significant behavioral differences among the three experimental groups in the first-stage period.

In the second stage (the period after the fourth practical class ($P4$)), the interventions are in place and used by the students. That is, each week students are asked to formulate a study goal for the upcoming week (all conditions) and to indicate via a detailed study plan how this goal will be reached (planner- and rabbit condition). This study plan, in turn, is used to construct the personalized pacesetter (rabbit condition). At first sight, the interventions do little in terms of inducing students to start earlier with studying. All conditions display the same pattern of student effort, which is similar to the pattern observed in the first stage. Most students continue to bunch their effort right before the due dates of graded tasks. I will later elaborate on the question to what extent this observed study behavior is consistent with the initial study plans of students.

Figure 6 examines whether the interventions increase student effort during the week. The figure plots per experimental condition the average weekly study time on the MicroApp, conditional on having logged in at least once during the week. Study time during the transition day between the first- and the second stage is excluded from the analysis. The figure shows that average student effort starts off low in the beginning of the course but gradually increases once exams are approaching. In the first stage (weeks 1-4), average student effort appears to be the same across the three conditions. In the second stage (weeks 5-8), students in the control condition (T1) and the rabbit condition (T3) continue to study at more or less the same pace. Students in the planner condition (T2), however, seem to lag behind in terms of study effort, especially in the exam week.

To formally identify the treatment effects δ on weekly study time, I adopt a differences-in-differences regression specification that models the MicroApp study time Y_{it} (in minutes) of student i in week t conditional on treatment status T_i and an indicator for the weeks with the interventions fully implemented (from week 5 on). I refer to this indicator

Figure 6: Average Weekly Study Time on the MicroApp



Notes: The figure shows per experimental condition the average weekly study time on the MicroApp, conditional on having logged in at least once during the week. Experimental variation is introduced in the in-class surveys during the fourth practical class (at the end of lecture week 4). Study time during this transition day is excluded from the analysis. Three different MicroApp versions (one per experimental condition) aired the next day (start of week 5) and stayed online until the final exam.

as $post_t$. Student fixed effects μ_i are also included to control for time-invariant student-specific characteristics that may affect study time, such as initial ability. The regression model is estimated with robust standard errors, clustered at the student level so as to account for within-student correlation patterns in the error term ϵ_{it} (Bertrand, Duflo and Mullainathan 2004). The model specification is as follows:

$$Y_{it} = \beta \cdot post_t + \sum_{j=1}^{3} I\{T_i = j\} \cdot (\delta_j \cdot post_t) + \mu_i + \epsilon_{it}. \tag{1}$$

Table 7 reports the results. The estimates in the table can be best interpreted as treatment-on-the-treated (ToT) effects, since students are only included when they logged in at least once during the week. My interest is mainly in the behavior of these students because they have been exposed to the treatment variation on the MicroApp and have at least attempted to exert non-zero study effort. Intention-to-treat (ITT) effects are presented in Appendix section A.7 and are qualitatively similar.[22]

[22]The ITT estimates are based on all students, including those with zero study effort on the MicroApp

Table 7: Treatment Effects on MicroApp Study Time

	(1)	(2)	(3)	(4)	(5)
Post=1	106.007***	119.549***	102.288***	109.408***	0.678***
	(23.141)	(21.338)	(22.082)	(19.850)	(0.151)
Post=1 × Treatment=2	-60.906**	-66.413**	-51.392*	-44.109*	-0.435*
	(30.906)	(28.164)	(29.477)	(26.635)	(0.239)
Post=1 × Treatment=3	-46.587	-32.771	-36.611	-25.937	-0.364
	(32.563)	(30.307)	(31.657)	(28.481)	(0.238)
Treatment=2		-1.165		2.950	
		(22.629)		(20.956)	
Treatment=3		22.294		21.872	
		(22.894)		(21.077)	
Constant	195.461***	180.656***	185.166***	169.808***	4.265***
	(6.288)	(15.189)	(6.362)	(14.098)	(0.049)
R^2	.0303	.0339	.0276	.03	.025
Number of students	305	305	413	413	305
Number of observations	1,174	1,174	1,376	1,376	1,174
Student fixed effects	Yes	No	Yes	No	Yes
Student inclusion/exclusion rules applied?	Yes	Yes	No	No	Yes
Time unit	Week	Week	Week	Week	Week
Dependent variable	Level	Level	Level	Level	Log

Notes: Identification of the treatment-on-the-treated effects on MicroApp study time (in minutes). Robust standard errors (clustered at the student level) are in parentheses. The variable *post* captures the experimental period. Study time during the transition day (5 October 2018) from the baseline period (lecture weeks 1-4) to the experimental period (lecture weeks 5-8) is excluded. There are two exclusion rules. First, a student is excluded when he or she participated in less than two practicals after the midterm exam. Second, a student is also excluded when he or she never made use of the MicroApp in the period until the first practical after the midterm exam. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Consistent with the interpretation of Figure 6, I find some (weak) support for the observation that students in the planner condition lag behind in terms of weekly study time on the MicroApp. Planner students study about one hour less per week compared to control students. This negative effect extends to the full sample of students (when no inclusion/exclusion rules are applied). Table 7 suggests that the pacesetter feature does not increase study time on top of what is already observed among control students in the experimental period. If anything, the pacesetter has a negative impact on study time as well, although these effects are not significantly different from zero at conventional significance levels.

With the available data, I therefore find no support for the first hypothesis in Section 3 in terms of overall treatment effects. Students who can observe their own pacesetter do not study more than students who cannot. Later I will examine study commitments more closely and discuss whether the treatments succeeded in helping students to reach their goals.

during the week. While the ITT point estimates of study time are lower in the different specifications, the results are qualitatively similar to the ToT estimates. As a robustness check, I also estimated the ITT effects with Tobit regressions and found similar results. These results are available upon request.

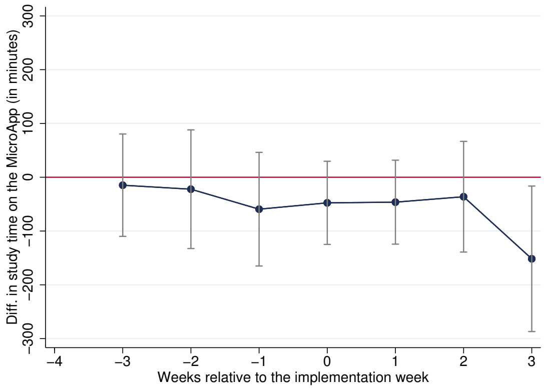
The absence of an overall treatment effect need not imply that there are no temporal effects of the interventions. Given the multi-week experimental period and the multiple rounds of setting study plans and pacesetters, this can be a real possibility. For example, the interventions may have had an impact only in the beginning of the experimental period. The richness of the data makes it possible to explore such temporal effects. Figure 7 shows the weekly treatment effects associated with the two interventions. The figure plots coefficients $\gamma_{j\tau}$ and $\delta_{j\tau}$ from the following regression specification:

$$Y_{i\tau} = \sum_{j=1}^{3} \sum_{\tau=m+1}^{-1} \gamma_{j\tau} D_{ij\tau} + \sum_{j=1}^{3} \sum_{\tau=0}^{q} \delta_{j\tau} D_{ij\tau} + \upsilon_{\tau} + \mu_i + \epsilon_{i\tau}. \quad (2)$$

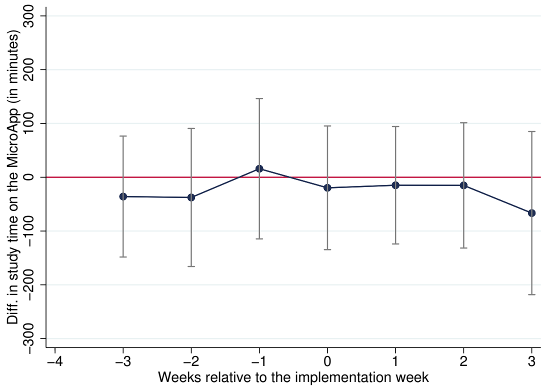
Where τ captures the proximity in weeks from the fifth lecture week (the implementation week), with $\tau \in [m, q]$. Note that m is the first lecture week ($\tau = -4$) and q is the exam week ($\tau = 3$). The implementation week (start of the experimental period) is $\tau = 0$. The leads (lags) are therefore captured by $\tau < 0$ ($\tau \geq 0$). Interactions between the treatment indicator T_i and the time dummies υ_τ are captured by $D_{ij\tau}$. As in Equation 1, the model is estimated with student fixed effects μ_i and robust standard errors (clustered at the student level). The first week is treated as the baseline.

The figure reveals no immediate effects of the treatments, nor do I in general observe delayed responses. At the end of the experimental period, close to the final exam, treated students appear to study slightly less than control students (at least on the MicroApp). In line with the results from Table 7, however, the overall impression is that the treatments do little in terms of inducing students to exert extra study effort. The figure also shows that the coefficients of the leads, $\gamma_{j\tau}$, are not significantly different from zero. This suggests pre-experimental balance and no anticipation effects.

Figure 7: Temporal Treatment Effects on MicroApp Study Time



(a) Condition II: Planner



(b) Condition III: Rabbit

Notes: The figure plots coefficients $\gamma_{j\tau}$ and $\delta_{j\tau}$ from Equation 2. Vertical spikes indicate 95% confidence intervals. The first lecture week is taken as the reference week. The fifth lecture week is the implementation week of the experiment (point 0 on the x-axis). The dependent variable is the weekly study time (in minutes) of a student on the MicroApp. Regressions are estimated with robust standard errors (clustered at the student level) and include student- and time fixed effects. Only students who have logged in at least once during the week are included in the sample. Study time during the transition day from the baseline period to the experimental period is excluded.

Further analyses indicate that there are no interaction effects of the treatments with gender status, pre-experimental course performance, and prior procrastination tendencies. In Table A.2, I separate the effects for males and females. The signs suggest that the pacesetter feature induces male students to study less compared to their counterparts in the control group. The negative effect is less pronounced among female students, but the difference with males is not significant. This pattern is consistently observed across the different model specifications. In Table A.3, I compare high-performance students with low-performance students (based on pre-experimental performance in the course) and observe that there is no treatment heterogeneity between these two groups. Students with procrastination tendencies also do not respond differently to the treatments compared to nonprocrastinating students, as is shown in Table A.4.

The previous analyses highlight that the two treatments are ineffective in increasing study effort among students. Importantly, though, this only concerns study effort on the MicroApp, as this is the learning input that was targeted by the treatments (for reasons discussed in Section 4.6). Thus far, I have not considered the possibility that the treatments affected student effort indirectly through effort in other learning inputs.

Figure A.10 in the appendix aggregates the in-class survey responses per lecture week in the experimental period.[23] There were no classes during the exam week and hence no survey responses are available for that week. The survey responses from the lecture weeks closely mirror the results from the baseline period. Most of the students' time is devoted to attending classes, followed by reading/practicing offline and use of the MicroApp. Other online resources appear to be rarely used by students.

In Table 8, I examine whether the treatments had an impact on time investments in the other learning inputs. The table shows the estimates based on interval regressions where the dependent variable is an interval measure of the time investment in a learning input. The weekly survey responses enable me to determine for each student the appropriate time interval that best represents his or her time investment in a learning input during a specific week. Furthermore, given that the in-class surveys were administered during both the baseline period and the experimental period, I can compare differences between

[23]Note that the second-stage survey responses show that a large share of students indicated that they spent on average less than one hour per week on the MicroApp. At first sight, this may not seem in line with Figure 6. The difference is that in Figure A.10 the responses are shown of all students, including the ones who never used the MicroApp during a specific week. Figure 6 only includes those students who used the MicroApp at least once during a specific week.

Table 8: Treatment Effects on Time Investments in Other Learning Inputs

Dependent variable:	(1) **Classes**	(2) **Offline Reading**	(3) **Offline Practice**	(4) **Online Read/watch**	(5) **Online Practice**	(6) **Other**	(7) **Other Courses**
Post=1	-0.738***	-0.250**	-0.503***	-0.123***	-0.121**	-0.182***	0.246
	(0.126)	(0.110)	(0.094)	(0.046)	(0.049)	(0.047)	(0.338)
Treatment=2	-0.140	-0.153	-0.161	-0.049	-0.077	-0.254***	-1.053*
	(0.242)	(0.191)	(0.138)	(0.083)	(0.066)	(0.074)	(0.557)
Treatment=3	-0.237	-0.200	-0.409***	0.064	-0.015	-0.156**	-1.451***
	(0.237)	(0.187)	(0.135)	(0.082)	(0.065)	(0.072)	(0.546)
Post=1 × Treatment=2	0.012	-0.042	-0.008	0.052	0.012	0.182***	0.383
	(0.177)	(0.156)	(0.133)	(0.065)	(0.069)	(0.067)	(0.481)
Post=1 × Treatment=3	-0.109	-0.382**	0.119	0.041	0.074	0.159**	0.432
	(0.174)	(0.153)	(0.131)	(0.065)	(0.069)	(0.067)	(0.467)
Constant	5.098***	2.678***	1.921***	0.726***	0.703***	0.812***	10.276***
	(0.171)	(0.133)	(0.096)	(0.058)	(0.046)	(0.051)	(0.390)
Number of students	309	309	309	309	309	309	308
Number of observations	1,926	1,922	1,920	1,923	1,926	1,921	1,614
Student inclusion/exclusion rules applied?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time unit	Week	Week	Week	Week	Week	Week	Week

Notes: Identification of the treatment effects on time investments in learning inputs other than the MicroApp. Estimates are based on interval regressions where the dependent variable is a time interval. At the end of each lecture week, students were asked through in-class surveys to report their weekly time investments in various learning inputs. The following four time intervals are used in the regressions: $[0,1)$, $[1,3)$,$[3,5)$,$[5,\infty)$, where the numbers are in hours. For the dependent variable *other courses* (average time investment in other courses), the following four time intervals are used: $[0,5)$, $[5,10)$,$[10,15)$,$[15,\infty)$, where the numbers are again in hours. The variable *post* captures the experimental period. No in-class surveys were administered at the end of the exam week. The learning input "classes" is partly incentivized (attendance is required to be eligible for a participation grade (worth 10% of the final grade)). The learning inputs "online read/watch" and "online practice" refer to non-MicroApp online study activities. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

the three conditions before and after experimental variation is introduced.

The table corroborates the previous finding that the treatments did not significantly alter study effort among students, nor did it lead to substitution between learning inputs. Prior to the experiment, students in the rabbit condition appear to engage in slightly less practicing offline and also invest less time in other courses. During the experimental period, however, I find no evidence that treated students engage differently with non-MicroApp learning inputs. Hence, the treatments seem to have had no impact on studying.

Study goals and goal completion rates Next, I discuss the treatment effects on study goal completion rates. The ex-ante reasoning was that both the planner and the rabbit condition would make progress toward a student's initial study goal more salient. The weekly study goal is transformed into a study plan with concrete daily goals that help the student to stay on track. These initial study commitments serve as reference points and induce the student not to postpone studying, thereby making the student less susceptible to self-control problems. I would thus expect that the interventions better enable students to reach their initial study goals.

Table 9: Percentage Shares of Non-Zero Weekly Study Goals

	T1: Control	**T2: Planner**				**T3: Rabbit**			
	%	%	Δ (T2-T1)	χ^2	p-value	%	Δ (T3-T1)	χ^2	p-value
All experimental weeks	80.27	82.30	2.03	0.49	0.485	83.51	3.24	1.31	0.252
Lecture week 5	82.98	87.91	4.93	0.90	0.342	88.89	5.91	1.40	0.237
Lecture week 6	80.41	81.72	1.31	0.05	0.818	78.43	-1.98	0.12	0.730
Lecture week 7	77.78	75.00	-2.78	0.21	0.648	79.38	1.60	0.07	0.784
Exam week	80.00	85.53	5.53	0.81	0.368	88.46	8.46	2.07	0.150

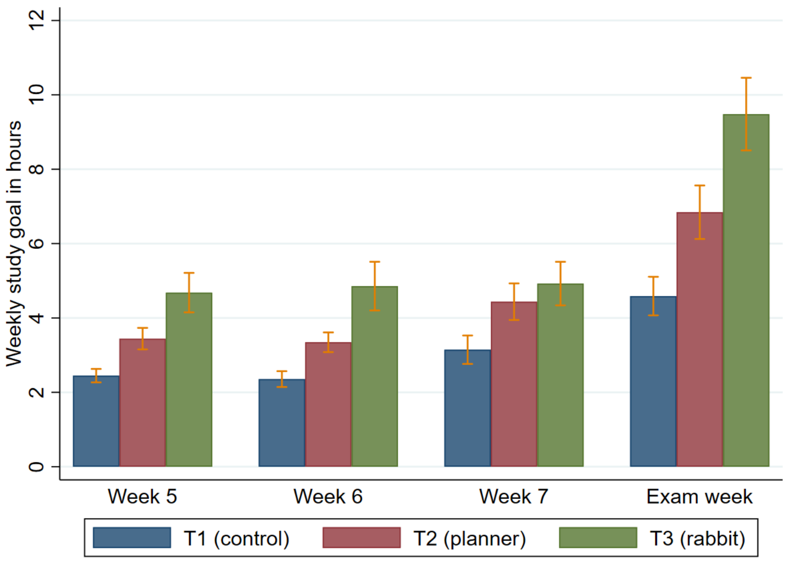
Notes: The table reports per experimental condition the percentage shares of students who have chosen a non-zero weekly study goal. The study goal is about how many hours students plan to study on the MicroApp in each week. Reported p-values are based on pairwise χ^2-tests between the control and treatment conditions.

Table 9 shows that most of the students choose to set non-zero weekly study goals (≈82%). The percentage shares are similar across the experimental conditions and do not vary much throughout the experimental period. I will restrict attention to these students with non-zero study goals, so as to avoid that goal completion rates are affected by students who trivially meet their goal by setting it equal to zero.[24]

[24] I also exclude weekly study goals that are unrealistically high (defined as five SDs above the mean). This leads to the exclusion of seven observations (0.78% of the sample). These students indicated that they plan to study more than 40 hours on the MicroApp in a week (which would be similar in hours to

A closer look at the weekly study goals reveals that students in both treatment conditions set significantly higher goals, as shown in Figure 8. Considering the full experimental period, pairwise Wilcoxon-Mann-Whitney (WMW) tests show that the differences with the control group are significant for both the planner treatment ($p < 0.001$) and the rabbit treatment ($p < 0.001$). The pacesetter, for example, induces students to set goals that are about twice as high as those set by students in the control group.

Figure 8: Weekly Study Goals for MicroApp Usage



Notes: The figure shows per experimental condition how many hours students on average plan to study on the MicroApp in each week, conditional on having set a non-zero study goal. Weekly study goals are elicited each week through in-class surveys. The error bars denote $\pm$ 1 standard error.

The pattern among treated students of setting higher study goals is observed for both male and female students. Figure A.11 reports the average study goal set by students in the experimental period, split by gender and treatment status. The figure shows clear within-gender differences between the different treatments. Both males and females in the planner and rabbit conditions set higher goals than their counterparts in the control condition.[25] Within treatments, I observe no significant differences between males and

a full-time lecture week). Not excluding these observations does not alter the main results: the p-values mentioned later in the main text are still $p < 0.001$ (based on pairwise WMW tests).

[25]Pairwise WMW tests yield the following p-values for male students: $p < 0.001$ (T2: planner) and $p < 0.001$ (T3: rabbit). For female students, the results are: $p = 0.002$ (T2: planner) and $p < 0.001$ (T3: rabbit).

females.[26] Hence, irrespective of gender, the treatments induce students to set more ambitious study goals, with no evidence that the effect is stronger or weaker for a particular gender group.

Similar results are obtained when a distinction is made between students who performed well or poor in the pre-experimental period of the course. Figure A.12 shows again that treated students set higher study goals, regardless of whether they performed well or poor before the interventions were in place.[27] Interestingly, only in the rabbit condition do I observe clear differences in goal-setting between high-performance students and low-performance students.[28] The latter group on average sets higher study goals (6.8 hours versus 4.8 hours). This difference is not observed in the planner condition, suggesting that it is the pacesetter feature that drives this result and not the act of formulating a study plan.

In Figure A.13, I show per experimental condition how procrastinating students differ from nonprocrastinating students in terms of their weekly study goals. As with gender and prior course performance, I observe once again that treated students set higher study goals, especially among those identified as nonprocrastinators.[29] Within the planner condition and the rabbit condition, the figure suggests that procrastinating students set lower study goals than nonprocrastinating students, although the effects are only marginally significant.[30]

A benefit of the MicroApp and the elicited study plans of students is that I can trace in real time how differences in planned and actual study time develop. In Figure 9, the green dot (red triangle) depicts planned (actual) daily study time on the MicroApp, conditional on having logged in at least once during the week and having set a non-zero weekly study goal. The bars show the daily difference between actual and planned study time, with a negative value indicating that students on average fell short of their own study goal

[26]The p-values based on WMW tests are: $p = 0.545$ (T1: control), $p = 0.795$ (T2: planner), $p = 0.955$ (T3: rabbit).

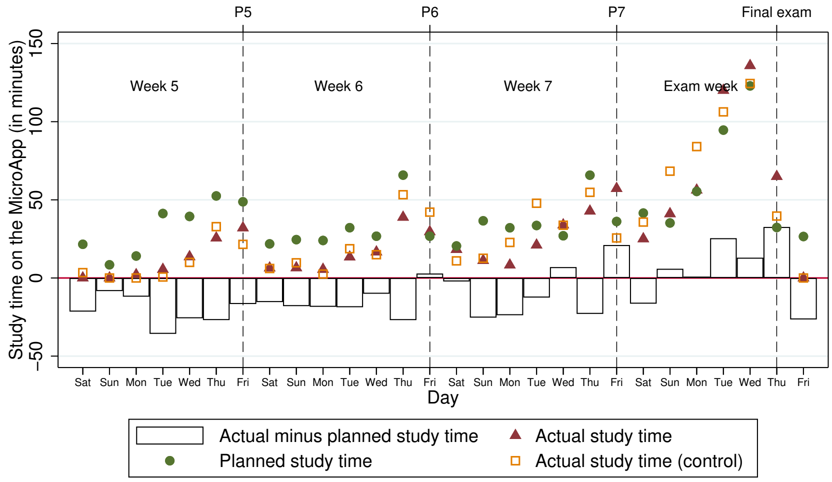
[27]Pairwise WMW tests yield the following p-values for high-performing students: $p < 0.001$ (T2: planner) and $p < 0.001$ (T3: rabbit). For low-performing students, the results are: $p = 0.009$ (T2: planner) and $p < 0.001$ (T3: rabbit).

[28]The p-values based on WMW tests are: $p = 0.028$ (T1: control), $p = 0.958$ (T2: planner) and $p < 0.001$ (T3: rabbit).

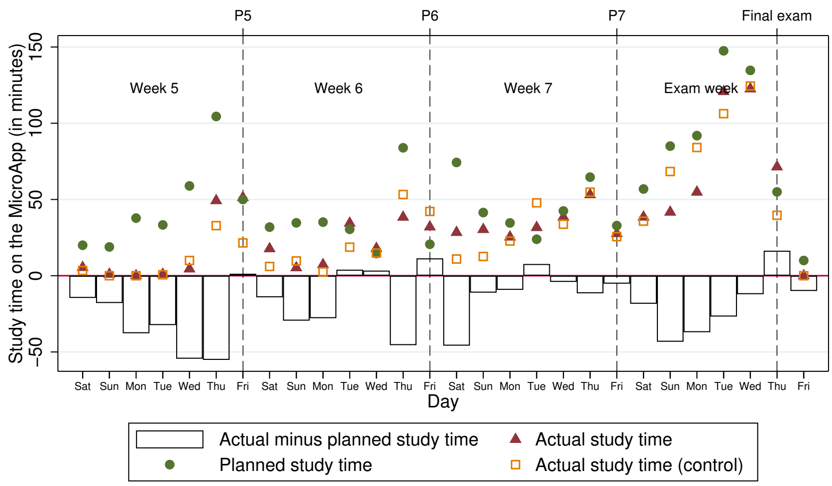
[29]Pairwise WMW tests yield the following p-values for nonprocrastinating students: $p < 0.001$ (T2: planner) and $p < 0.001$ (T3: rabbit). For procrastinating students, the results are: $p = 0.001$ (T2: planner) and $p < 0.001$ (T3: rabbit).

[30]The p-values based on WMW tests are: $p = 0.580$ (T1: control), $p = 0.055$ (T2: planner) and $p = 0.046$ (T3: rabbit).

Figure 9: Actual Versus Planned Study Time on the MicroApp



(a) Condition II: Planner



(b) Condition III: Rabbit

Notes: The figure shows per experimental condition the daily actual (red triangle) and planned (green dot) study time on the MicroApp during the second stage of the study (when experimental variation was implemented), conditional on having logged in at least once during the week and having set a non-zero weekly study goal. The bars shows the daily difference between actual and planned study time, with a negative value indicating that students on average fell short of their preset goal for that day. The hollow orange square shows the average daily actual study time of students in the control group (condition I). A graded quiz is administered during each practical class (*P*) on Friday.

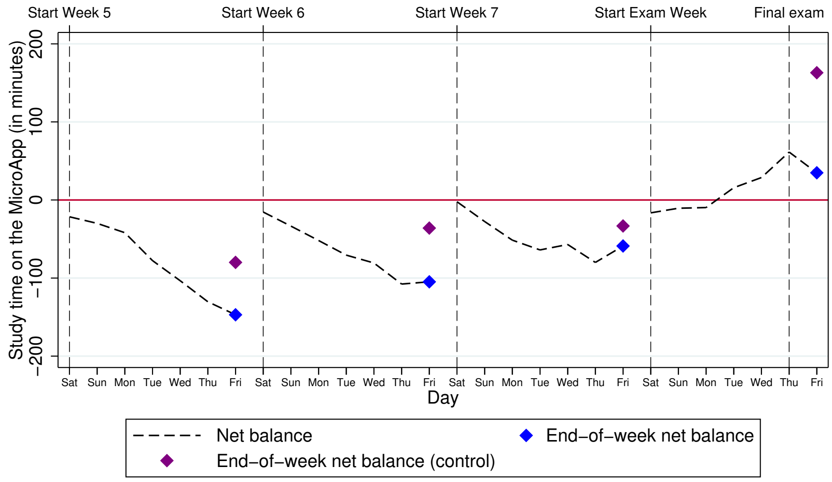
of that day. In general, students in both treatment conditions are too optimistic when planning their daily study effort; planned study time is in most cases higher than actual study time. The figure also shows that both planned and actual effort increase in the days before the practical quizzes and the final exam, which suggests that effort bunching is in part planned beforehand. The overall impression from the figure is that students seem to have difficulties in sticking to their initial plans.

I explore this further in Figure 10. The figure shows the net balance of the daily differences between actual and planned study time. A negative end-of-week net balance indicates that students exerted lower effort by the end of the week than initially planned. I observe that in most cases the net balance is negative and growing in size during the week. This is in contrast to what one would expect to see when students catch up. In case of catching up, the negative net balance should become smaller (or turn positive) later on in the week (when the weekly study goal is due). The figure reveals that this is not the case among treated students. Furthermore, I not only observe that treated students have negative end-of-week net balances, but that these negative balances are also larger than in the control group. Earlier, I found no evidence that treated students study more or less on the MicroApp, which implies that the larger negative end-of-week net balances are to a large extent due to having set more ambitious study goals.

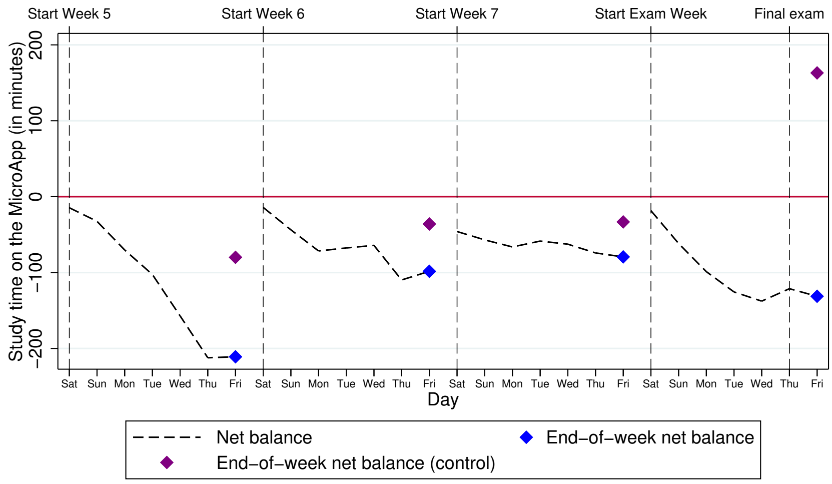
The previous figures provided some first visual evidence that treated students have difficulties in completing their weekly study goals. In Table 10, I contrast the performance of treated students with that of control students and identify the treatment effects on study goal completion rates. The table reports the logit marginal effects on the likelihood of achieving the study goal in a given week, conditional on having set a non-zero goal. The dependent variable is an indicator variable equal to one if student i achieved his or her study goal in week t, zero otherwise. The time period is from the fifth lecture week (start of the experimental period) until the exam week (end of the experimental period).

In column (1) of the table I find some initial, albeit weak, evidence that treated students are indeed *less* likely to reach their study goal by the end of the week, even after controlling for goal levels. In column (2), I restrict attention to those students who logged in once or more on the MicroApp during the week. These students have at least attempted to reach their goal, and, importantly, were exposed to the experimental variation on the MicroApp. Among these students, the negative effect persists and is larger

Figure 10: Net Balance Actual Vs. Planned Study Time on the MicroApp



(a) Condition II: Planner



(b) Condition III: Rabbit

Notes: The dashed black line shows per experimental condition the net balance per day between the actual and planned study time on the MicroApp, conditional on having logged in at least once during the week and having set a non-zero weekly study goal. A negative value indicates that students are behind their own schedule for reaching the study goal. The study goal is set each week on Friday during an in-class survey and is due next week's Friday 23:59. The purple diamonds indicate the end-of-week net balance of students in the control group. Control students only submitted a weekly study goal, and not a daily study plan.

in size: students in the planner condition and the rabbit condition are respectively about 13 and 14 percentage points less likely to reach their goal by the end of the week. When I do not control for the size of the study goal, as in column (3), these numbers change to 16 and 20 percentage points, respectively. This suggests that part of the negative effect is explained by treated students setting more ambitious study goals (as identified earlier). Similar results are obtained when the full sample of students is considered (that is, when no inclusion/exclusion rules are applied). I therefore find no support for hypothesis II in Section 3. To the contrary, I find some evidence that treated students are less likely to reach their study goals by the end of the week, partly as a result of setting more ambitious study goals than students in the control group.

Table 10: Treatment Effects on Study Goal Completion (Logit MEs)

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment=2	-0.045	-0.126**	-0.155**	-0.024	-0.095*	-0.119**
	(0.037)	(0.059)	(0.063)	(0.030)	(0.055)	(0.058)
Treatment=3	-0.079**	-0.139**	-0.202***	-0.053*	-0.110*	-0.164***
	(0.038)	(0.061)	(0.060)	(0.031)	(0.058)	(0.058)
Weekly study goal	-0.005	-0.026***		-0.002	-0.025***	
	(0.003)	(0.005)		(0.002)	(0.005)	
Number of students	283	211	211	403	251	251
Number of observations	875	432	432	1,194	496	496
Student inclusion/exclusion rules applied?	Yes	Yes	Yes	No	No	No
Conditional on weekly effort > 0	No	Yes	Yes	No	Yes	Yes

Notes: The table reports the treatment effects on the likelihood of achieving the study goal, conditional on having set a non-zero study goal. The time period under consideration is the experimental period (lecture week 5 - exam week). The dependent variable is an indicator variable equal to one if a student completed his or her study goal for the week, zero otherwise. Reported coefficients are logit marginal effects with robust standard errors in parentheses. The variable *study goal* is the study goal (in hours) set by student i in week t. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

In the appendix, I examine the treatment effects in more detail by exploring gender differences, differences between students identified as either high-performing or low-performing based on pre-experimental data, and differences between procrastinating students and nonprocrastinators. Figure A.14 shows the gender-specific treatment effects. Previous studies have shown that goal-setting is more effective for males than for females (e.g., Clark et al. (2017)). While the point estimates in the figure suggest that the adverse treatment effects are indeed mainly observed among female students, the difference with male students is not significant.[31] Figure A.15 shows separate treatment effects for low-performing students and high-performing students. Among these two groups, I observe no

[31] Adding a gender dummy in the specification of column (2) in Table 10 and interacting this dummy with the treatment indicators yields the following p-values associated with the interaction effects: $p = 0.218$ (T2: planner) and $p = 0.618$ (T3: rabbit).

differences in the treatment effects related to goal completion.[32] Figure A.16 distinguishes between procrastinators and nonprocrastinators. I find some evidence that nonprocrastinators in the planner group ($p = 0.037$) and the rabbit group ($p = 0.041$) are less likely to reach their goals compared to their counterparts in the control group. The treatment differences in goal completion rates between procrastinators and nonprocrastinators are not significantly different from zero.[33]

Do the study goal completion rates of treated students improve over time during the experimental period? Each week, students observe their own goal progress and know by the end of the week whether they succeeded in completing the goal. In case of no success, students may be prompted to adjust their goal to more realistic levels. The act of formulating a study plan and observing one's own pacesetter may help in this process. Figure 11 shows the goal completion rates per week and experimental condition. Completion rates do improve over time, suggesting that students set more realistic goals. However, consistent with the previous results, the figure provides no support for the claim that these improvements are due to the treatment interventions, as the improvements are also observed among control students. In fact, even though all three conditions have similar low completion rates at the start of the experimental period, control students consistently achieve higher completion rates in subsequent weeks, giving rise to the negative treatment effects observed in Table 10.

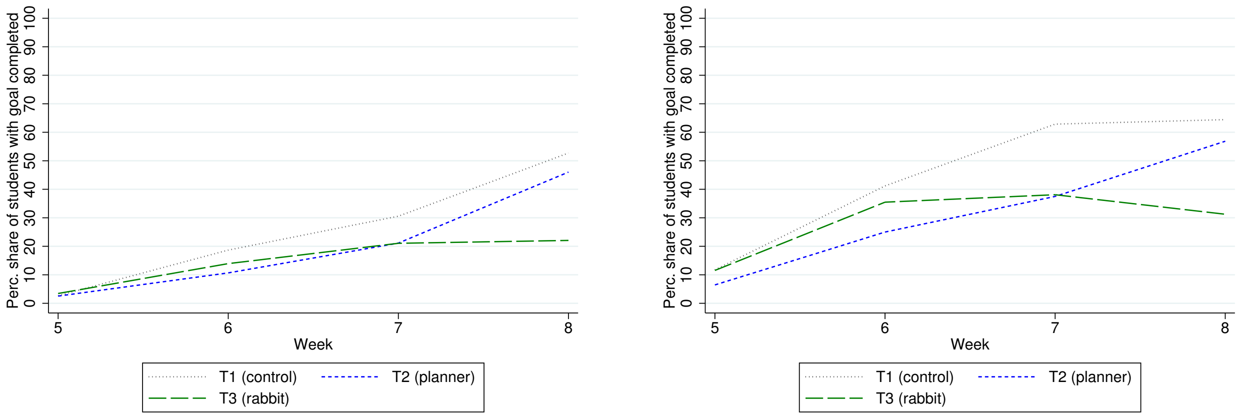
Learning outcomes In Section 3 it was hypothesized that student effort serves as the mediator between the treatment interventions and learning outcomes. Insignificant treatment effects on student effort means that the supposed channel through which I expected improved learning outcomes is absent. This makes it unlikely to find support for the third hypothesis, which states that students who can observe their own pacesetter will have better learning outcomes than students who cannot, conditional on the pacesetter being successful in increasing student effort.

Figure 12 shows the learning outcomes per experimental condition. The final course

[32] Adding a low-performance dummy in the specification of column (2) in Table 10 and interacting this dummy with the treatment indicators yields the following p-values associated with the interaction effects: $p = 0.696$ (T2: planner) and $p = 0.638$ (T3: rabbit).

[33] Adding a procrastination dummy in the specification of column (2) in Table 10 and interacting this dummy with the treatment indicators yields the following p-values associated with the interaction effects: $p = 0.294$ (T2: planner) and $p = 0.543$ (T3: rabbit).

Figure 11: Study Goal Completion Rates During the Experimental Period



(a) Cond. on study goal > 0

(b) Plus cond. on study time > 0

Notes: Study goal completion rates during the experimental period. Weekly completion rates are shown for each treatment condition separately. Figure (a) shows the rates for all students who submitted a non-zero study goal for the week. Figure (b) only considers those students who submitted a non-zero goal and additionally exerted non-zero study time on the MicroApp during the week.

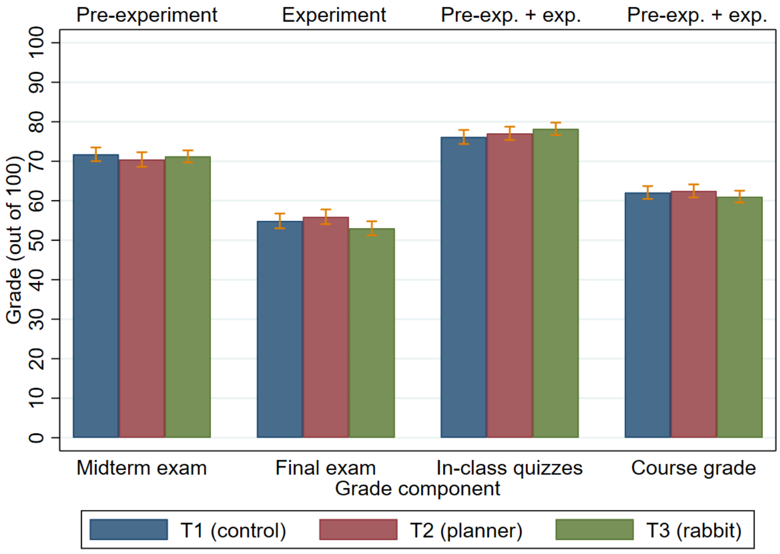
grade is shown as well as the performance on each course grade component separately. A convenient feature of the research design is that I can compare learning outcomes before and after experimental variation is introduced. For instance, the midterm exam is administered before the start of the experiment and we should therefore not expect to see significant midterm performance differences between the experimental conditions. The figure shows that midterm scores are indeed balanced across the three groups.[34] The final exam is administered at the end of the experiment and focuses on course material covered during the experimental period. This makes final exam performance a suitable outcome variable for measuring the treatment effects on learning outcomes, as I also indicated in the pre-analysis plan of this study. In the figure, I observe no evidence that treated students performed better on the final exam than control students.[35] This is in line with the previous reasoning that, absent increased student effort as the mediator, we should not expect to find a treatment effect on learning outcomes. Similar performance patterns are observed when looking at in-class quizzes and the final course grade of students.

[34] Pairwise comparisons with the control group yield the following p-values (two-sided t-tests): $p = 0.606$ (T2: planner) and $p = 0.822$ (T3: rabbit).

[35] Pairwise comparisons with the control group yield the following p-values (two-sided t-tests): $p = 0.812$ (T2: planner) and $p = 0.187$ (T3: rabbit).

The treatment effects on learning outcomes (final exam performance) are formally identified in Table 11. Looking at students who satisfy the inclusion rules of this study (columns (1) and (2)), I observe that the point estimates are small in size and not significantly different from zero. In analyses not shown for brevity I also looked at the final course grade as a measure of learning outcomes and found very similar results.

Figure 12: Learning Outcomes per Experimental Condition



Notes: The figure shows the learning outcomes per experimental condition. For ease of comparison, all grades are converted to be on a 0-100 scale, with a higher score indicating better performance. The weight of each course grade component is as follows: midterm exam (30%), final exam (60%), and in-class quizzes (10%). The midterm exam (final exam) was administered before (after) the introduction of experimental variation. The final exam focused on material covered during the experimental period. Performance on in-class quizzes is the average of a student's five best quizzes (out of six). Quizzes were administered every lecture week during the (pre-)experimental period. The error bars denote $\pm$ 1 standard error.

In the appendix, I present the treatment effects for each of the three subgroups discussed in this chapter. Table A.5 shows the results separately for males and females. Female students scored on average about seven points higher on the final exam than male students. Interacting treatment- and gender indicators yields some (weak) evidence that the pacesetter feature triggered an adverse effect among female students. Compared to male students in the rabbit condition, female students who observed their own pacesetter scored an additional nine points lower on the final exam. Given that this result is only marginally significant (at the 10%), I view this as suggestive rather than conclusive evidence that the pacesetter caused the adverse effect among females. In Table A.6 and Table A.7, I respectively consider whether treatment effects differ for students based on their pre-experimental course performance or prior procrastination tendencies. The estimates show that this does not seem to be the case.

Table 11: Treatment Effects on Learning Outcomes (Final Exam Score)

	(1)	(2)	(3)	(4)
Treatment=2	-0.570	-0.475	-0.862	-0.797
	(2.390)	(2.381)	(1.845)	(1.836)
Treatment=3	-3.045	-2.979	-4.198**	-4.149**
	(2.299)	(2.306)	(1.860)	(1.859)
Constant	57.083***	56.248***	55.154***	54.060***
	(1.592)	(1.820)	(1.289)	(1.385)
Number of students	302	302	504	504
Student inclusion/exclusion rules applied?	Yes	Yes	No	No
Control for second-stage effort on the MicroApp?	No	Yes	No	Yes

Notes: The table reports the treatment effects on learning outcomes. The dependent variable is the number of points scored on the final exam (out of 100 points). Second-stage effort on the MicroApp refers to the total study time of a student on the MicroApp during the experimental period. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

Table 12: Treatment Effects on Achieving Course Goals (Logit MEs)

Dependent variable:	(1) **Midterm Exam**	(2) **Final Exam**	(3) **Course Grade**
Treatment=2	-0.023	-0.015	-0.053
	(0.066)	(0.042)	(0.047)
Treatment=3	-0.025	0.045	0.044
	(0.065)	(0.046)	(0.053)
Number of students	309	309	309
Student inclusion/exclusion rules applied?	Yes	Yes	Yes

Notes: The table reports the treatment effects on the likelihood of achieving the initially formulated course goals. At the start of the course, each student was asked in an in-class survey to indicate their grade target for the midterm exam, the final exam, and the overall course. The dependent variable is equal to one if the actual grade is equal or higher than this target grade, zero otherwise. Reported coefficients are logit marginal effects. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

One concern with using increases in final exam scores as a proxy for improvements in learning outcomes is that it does not account for what students themselves wanted to achieve in the course. When the marginal cost of exerting extra effort exceeds the marginal benefit, a student may not find it worthwhile to go for a higher grade. In support of this, students indicated at the start of the course that they do not seek a perfect score on the exams but rather aim for a grade of around 7.5-8 (out of 10), as reported in Table 4. Therefore, in Table 12 I alternatively look at whether the treatments better enable students to reach their initial course goals. The dependent variable is equal to one if the actual grade is equal or higher than the initial target grade of the student, zero otherwise. I observe that the treatments do not increase the likelihood of achieving initial goals.

5.3 Additional analyses

Correlations between learning inputs and learning outputs One concern with targeting interventions at a particular learning input is that it may divert the attention of students from other, potentially more productive, learning inputs. In the previous analyses, I found no evidence that the treatments altered the way students studied in the course. No between-treatment differences were observed in terms of time investments in learning inputs. From a policy perspective, however, the question which learning inputs contribute to improved academic outcomes, and thus which inputs should be targeted, continues to be an important research topic that should inform the design and implementation of future educational interventions. With the collected data in this study, I can perform an exploratory analysis that serves as a first step in this direction.

The learning input that was targeted by the interventions in this study was usage of the MicroApp, an online learning platform with ample practice- and recap material. Motivating students to use a particular learning input typically follows from the belief that the input improves learning outcomes. Since I did not randomize access to the MicroApp, I can not infer the causal impact the platform has had on student outcomes, and I can therefore not say to what extent this belief is warranted. What I can do with the available data is to look whether performance on the MicroApp correlates positively with performance in the course. Finding such a correlation is insightful for two reasons. One reason is that it would suggest that interim performance on the MicroApp has predictive power with regard to future learning outcomes, which can be of great value to researchers and educators who want to use educational technologies to identify, for example, at-risk students in real time. The second reason is that a positive correlation suggests that the relation between input and output goes in the expected direction.

In Figure A.17, I plot the final course grade of students against their performance on the MicroApp. I restrict attention to those students who have solved at least 100 tasks on the app and measure performance as the percentage of correctly solved MicroApp practice tasks. The fitted regression line is upward sloping, indicating a positive correlation between performance in the course and on the MicroApp. The predicted values are intuitive: a student who has solved, say, 60% of MicroApp tasks correctly is expected to complete the course with a score of about six out of ten.

What is the productivity of each learning input in terms of improving learning outcomes? In Figure A.18, I provide the results of an exploratory analysis that takes the following approach. For each lecture week, I use the in-class survey responses to determine for every student his or her time investment in a particular learning input. I then relate these time investments to the performance on in-class quizzes, covering the course material of that week. Specifically, I regress the quiz grade on student fixed effects and on each level of categorical variables that capture students' weekly time investment in a learning input. The reference category for an input consists of students with no time spent on that input during a specific week. For most inputs, there is no significant difference in quiz grades when more time is devoted to that input. Interestingly, I do find some evidence that quiz grades improve when more time is spent on the MicroApp, which suggests that the treatments were targeted at the right input.

6 Discussion

Technological innovations in education yield widespread excitement among policy makers. Research on these innovations, however, has not kept up with the pace at which these innovations are proliferating in educational programs, as noted by Escueta et al. (forthcoming). Most of the existing research has focused on comparing courses with digital learning components to more traditional courses with offline learning components. Swoboda and Feiler (2016), for example, find that student performance improves in courses that employ technological tools, such as online homework assignments and video lectures. Mixed results are reported in other studies.[36] More research is therefore needed that ex-

[36]Lee et al. (2010) find that students have favorable attitudes toward digital learning tools, and 73% of the students prefer these digital tools over traditional homework. Though, they find no evidence that online homework is more effective in improving test scores than traditional homework. Sosin et al. (2004) find a small but positive effect on student performance in classes that use technology. Hernandez-Julian and Peters (2012) find that students who submit homework online complete more assignments but have lower class attendance rates. They find no effect on exam scores when homework is completed online. Brown and Leidholm (2002) find that students in virtual classes perform significantly worse on exams

plores in detail how students interact with these technologies and how technologies should be designed to optimally improve the learning outcomes of students.

This chapter aimed to achieve just that. I developed a new educational technology to meticulously measure student effort, to contrast it with other learning inputs, and to administer personalized planning interventions to students in real time. I implemented a field experiment and used A/B-testing within the technology to identify the causal effects of different design features on student effort and outcomes. The empirical results shed light on how students interact with educational technologies, how their effort evolves throughout the course, and how outcomes are affected by technological and non-technological learning inputs. Subgroup analyses based on gender, ability and procrastination tendencies have further facilitated an understanding of these results.

The finding that the pacesetter feature does not alter effort and outcomes is highly relevant for the literature on commitment mechanisms in education. Most existing commitment mechanisms, such as deadlines and goals, tend to be static and may not be based on students' own preferred study plans. The pacesetter, instead, dynamically links different selves by letting students "race" against their own initially preferred study pace. The pacesetter is thus fully consistent with a student's own initial plan. Furthermore, the pacesetter preserves flexibility, but still triggers the motivating effect of loss aversion when falling behind. While my analysis shows that students actively engage with their pacesetter, by setting non-zero study goals that are more ambitious than those set by control students, I find no evidence that this materializes into increased study effort. More research is needed that explores which type of pacesetter is most effective in bringing about changes in student effort and learning outcomes.

Through weekly surveys at the student level and MicroApp activity trackers, I have acquired detailed data on students' time investments in multiple learning inputs. This information is often limited or even absent in existing studies on student effort, forcing these studies to proxy effort by focusing on outcome measures instead. The data in this study yield rich insights on how students actually study. I observe that students invest most of their time in attending classes, reading offline course material, and practicing with offline material. With the exception of the MicroApp, students make surprisingly little use of other online resources. When exams are approaching, students shift most of

than students in a live course.

their attention from "consuming" the course material to practicing. Comparing planned and actual study effort over time, I showed that students exert lower-than-planned effort right from the start and on average do not catch up later on.

In this study, I explored treatment heterogeneity in detail in order to see which specific groups benefit the most from the interventions. Duckworth et al. (2015) conjecture that educational interventions aimed at improving self-control may be especially beneficial for male students. I find no evidence of this for the specific interventions in this study. Male students in the treatment conditions do not study more, nor do they have improved learning outcomes. The absence of an effect is in contrast with Clark et al. (2017), who do find that task-based goals are more effective for male students than for female students. Considering treatment heterogeneity based on procrastination, Himmler et al. (2019) find that commitment devices in education are most effective for procrastinating students. I find no evidence of this in the present field experiment. There is also no treatment heterogeneity when students are compared based on initial ability (pre-experimental performance in the course). Previous research has shown that incentives improve (worsen) the educational achievements of high-ability students (low-ability students) (Leuven, Oosterbeek and van der Klaauw 2010). I find no treatment differences between low-ability students and high-ability students in a setting with non-financial incentives.

7 Conclusion

A lack of study time often prevents students from reaching their academic potential. While recent studies suggest that motivating students to study may be better achieved when effort is incentivized rather than output, only limited progress on this is made thus far due to a lack of data on effort and initial study plans. Procrastination tendencies further exacerbate the challenge of getting students to study more. These issues have stalled progress on designing and evaluating interventions that target student effort.

In a field experiment with 573 university students, I elicit detailed weekly study goals and plans from all students and use these to construct individualized pacesetters (*rabbits*). Pacesetters are moving reference points that visualize the preferred study pace of the present self by moving exactly according to the initial study plan. I build a new educational technology to measure student effort and to present the pacesetters to students in real time. Students can at any time see how they perform relative to their own pacesetter. Falling behind immediately makes salient to the student that he or she is reneging on earlier study commitments. The pacesetter thus serves as an internal commitment mechanism whereby future selves are triggered to follow the pacesetter and, in doing so, implement the initial study plan.

The experiment is designed to identify the causal effect of the pacesetter on student effort and learning outcomes, and to control for confounding effects related to the acts of formulating study goals and plans. The findings are as follows. First, I find that the pacesetter induces students to set more ambitious study goals, amounting to about two additional study hours per week. This, however, does not lead to more effort exerted. I find that the pacesetter has no impact on study time, as measured by the educational technology, nor do I observe effort spillovers to other learning inputs and other courses. The richness of the data allows me to examine exactly when actual study time deviates from planned study time. I observe that students already fall behind their pacesetter at the start of the week and on average show no tendency to catch up later in the week. I consider multiple measures of learning outcomes and find no effect on these outcomes.

In additional analyses, I consider treatment heterogeneity based on gender, ability, and prior procrastination tendencies. In terms of student effort, I find no heterogeneity within these groups. With regard to goal-setting behavior and goal completion rates, I find some evidence that the pacesetter induces low-ability students (procrastinating students) to set higher (lower) weekly study goals. In all three groups, however, no differences are observed in the likelihood of actually achieving the weekly study goals. There are also no within-group differences observed for learning outcomes, with the exception of suggestive evidence that the pacesetter caused females to perform poorer on the exam.

This study is one of the first attempts to use educational technologies as a means to target behaviors that have a negative impact on academic performance. I show how these technologies can be used to measure student effort in real time, to administer personalized interventions, and to implement experimental evaluations through A/B-testing. In a review of the literature, Escueta et al. (forthcoming) conclude that technology-enabled interventions have great potential for improving education-related outcomes, but they also observe that existing studies have mostly limited their attention to campaigns with text messages. I show instead the potential of technologies to administer fully customized data-driven interventions at low marginal cost.

The present study points to several directions for future research. First, more studies are needed that explore how technology can be used to target behaviors that adversely affect academic performance. Such studies will be particularly fruitful when they utilize the unique benefits that technology has to offer, from meticulous measurement of student behavior to implementing personalized interventions. While this study has focused on procrastination and self-control problems among students, I reckon that other behaviors are suitable for further investigation as well. Second, this study has used students' own initial study plans as input for the construction of the pacesetter. Interesting possibilities for future research are to use alternative sources that determine the pace of the pacesetter.

The digitalization of education unveils many new opportunities and challenges for researchers and educators alike. The ease and frequency with which students use technology in their everyday lives likely results in that studying will increasingly take place online. How students learn in a digital world is an important question with profound implications for how human capital will be accumulated in the future.

References

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutströmd, "Discounting behavior: A reconsideration," *European Economic Review*, October 2014, *71*, 15–33.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review*, December 2012, *102* (7), 3333–3356.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos, "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 2009, *1* (1), 136–163.

Ariely, Dan and Klaus Wertenbroch, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 2002, *13* (3), 219–224.

Ashraf, Nava, Colin F. Camerer, and George Loewenstein, "Adam Smith, Behavioral Economist," *Journal of Economic Perspectives*, 2005, *19* (3), 131–145.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over Time: Dynamic Inconsistency in Real Effort Tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

Babcock, Philip and Mindy Marks, "The Falling Time Cost of College: Evidence form Half a Century of Time Use Data," *The Review of Economics and Statistics*, 2011, *93* (2), 468–478.

Benabou, Roland and Marek Pycia, "Dynamic inconsistency and self-control: a planner-doer interpretation," *Economics Letters*, 2002, *77*, 419–424.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates," *The Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Brown, Byron W. and Carl E. Liedholm, "Can Web Courses Replace the Classroom in Principles of Microeconomics?," *American Economic Review*, May 2002, *92* (2), 444–448.

Brown, Jeffrey R. and Alessandro Previtero, "Procrastination, Present-Biased Preferences, and Financial Behaviors," *NBER Working Paper Series*, 2014.

Bryan, Gharad, Dean Karlan, and Scott Nelson, "Commitment Devices," *Annual Review of Economics*, 2010, *2*, 671–698.

Burger, Nicholas, Gary Charness, and John Lynham, "Field and online experiments on self-control," *Journal of Economic Behavior & Organization*, March 2011, *77* (3), 393–404.

Carvalho, Leandro S., Stephan Meier, and Stephanie W. Wang, "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday," *American Economic Review*, 2016, *106* (2), 260–284.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush, "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments," *NBER Working Paper Series*, 2017, (WP No. 23638).

Duckworth, Angela L., Elizabeth P. Shulman, Andrew J. Mastronarde, Sarah D. Patrick, Jinghui Zhang, and Jeremy Druckman, "Will Not Want: Self-Control Rather than Motivation Explains the Female Advantage in Report Card Grades," *Learning and Individual Differences*, 2015, *39*, 13–23.

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos, and Vincent Quan, "Upgrading Education with Technology: Insights from Experimental Research," *Journal of Economic Literature*, forthcoming.

Frederick, Shane, *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, New York: Russell Sage Foundation,

Giné, Xavier, Jessica Goldberg, Dan Silverman, and Dean Yang, "Revising Commitments: Field Evidence on the Adjustment of Prior Choices," *The Economic Journal*, February 2018, *128* (608), 159–188.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, December 2004, *42* (4), 1009–1055.

Heffetz, Ori, Ted O'Donoghue, and Henry S. Schneider, "Forgetting and Heterogeneity in Task Delay: Evidence from New York City Parking-Ticket Recipients," *NBER Working Paper Series*, 2016, (WP No. 23012).

Hernandez-Julian, Rey and Christina Peters, "Does the Medium Matter? Online versus Paper Coursework," *Southern Economic Journal*, April 2012, *78* (4), 1333–1345.

Hershfield, Hal E., "Future self-continuity: how conceptions of the future self transform intertemporal choice," *Annals of the New York Academy of Sciences*, October 2011, (1235), 30–43.

___, Daniel G. Goldstein, William F. Sharpe, Jesse Fox, Leo Yeykelis, Laura L. Carstensen, and Jeremy N. Bailenson, "Increasing saving behavior through age-progressed renderings of the future self," *Journal of Marketing Research*, November 2011, *48*, 23–37.

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk, "Soft Commitments, Reminders, and Academic Performance," *American Economic Journal: Applied Economics*, 2019, *11* (2), 114–142.

Hsiaw, Alice, "Goal Bracketing and Self-Control," *Working paper*, 2015.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen, "Behavioral economics of education," *Journal of Economic Behavior & Organization*, 2015, *115*, 3–17.

Koch, Alexander K. and Julia Nafziger, "Goals and bracketing under mental accounting," *Journal of Economic Theory*, March 2016, *162*, 305–351.

___ and ___, "Motivational Goal Bracketing: An Experiment," *Working paper*, 2017.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, May 1997, *112* (2), 443–478.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos, "Behavioral economics of education: progress and possibilities," *Handbook of the Economics of Education*, 2016, *5*, 1–74.

Lee, William, Richard H. Courtney, and Steven J. Balassi, "Do Online Homework Tools Improve Student Results in Principles of Microeconomics Courses?," *American Economic Review*, May 2010, *100* (2), 283–286.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," *Journal of the European Economic Association*, 2010, *8* (6), 1243–1265.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff, "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance," *American Economic Journal: Economic Policy*, November 2016, *8* (4), 183–219.

Martinez, Seung-Keun, Stephan Meier, and Charles Sprenger, "Procrastination in the Field: Evidence from Tax Filing," *Working paper*, 2017.

Meier, Stephan and Charles D. Sprenger, "Temporal Stability of Time Preferences," *The Review of Economics and Statistics*, May 2015, *97* (2), 273–286.

Metcalfe, Robert, Simon Burgess, and Steven Proud, "Students' effort and educational achievement: Using the timing of the World Cup to vary the value of leisure," *Journal of Public Economics*, April 2019, *172*, 1111–1126.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, March 1999, *89* (1), 103–124.

OECD, "Education at a Glance 2008: OECD Indicators," Technical Report, OECD 2008.

Oreopoulos, Philip, Richard W. Patterson, Uros Petronijevic, and Nolan G. Pope, "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*, 2018, (WP No. 25036).

Paola, Maria De and Vincenzo Scoppa, "Procrastination, Academic Success and the Effectiveness ofa Remedial Program," *Journal of Economic Behavior & Organization*, July 2015, *115*, 217–236.

Reuben, Ernesto, Paolo Sapienza, and Luigi Zingales, "Procrastination and Impatience," *Journal of Behavioral and Experimental Economics*, October 2015, *58*, 63–76.

Sosin, Kim, Betty J. Lecha, Rajshree Agarwal, Robin L. Bartlett, and Joseph I. Daniel, "Efficiency in the Use of Technology in Economic Education: Some Preliminary Results," *American Economic Review*, May 2004, *94* (2), 253–258.

Steel, Piers, "The Nature of Procrastination: A Meta-Analytic and Theoretical Review of Quintessential Self-Regulatory Failure," *Psychological Bulletin*, 2007, *133* (1), 65–94.

Stinebrickner, Ralph and Todd R. Stinebrickner, "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*, June 2008, *8* (1), 1–55.

Swoboda, Aaron and Lauren Feiler, "Measuring the Effect of Blended Learning: Evidence from a Selective Liberal Arts College," *American Economic Review*, May 2016, *106* (5), 368–372.

Thaler, Richard H and Hersh M. Shefrin, "An Economic Theory of Self-Control," *Journal of Political Economy*, February 1981, *89* (2), 392–406.

Wong, Wei-Kang, "How much time-inconsistency is there and does it matter? Evidence on self-awareness, size, and effects," *Journal of Economic Behavior & Organization*, December 2008, *68* (3-4), 645–656.

A Appendix

A.1 Screenshots of the MicroApp (pre-experiment)

Figure A.1: Dashboard With Basic Learning Analytics

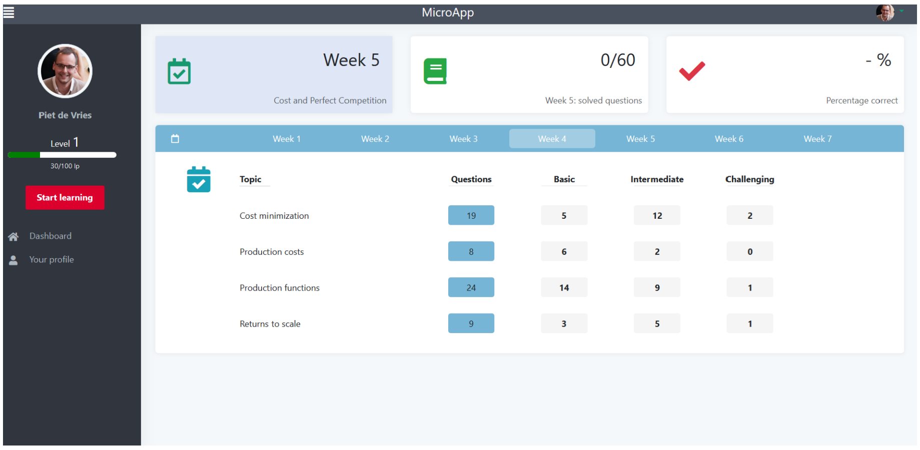


Figure A.2: Answered Practice Task (With Feedback)

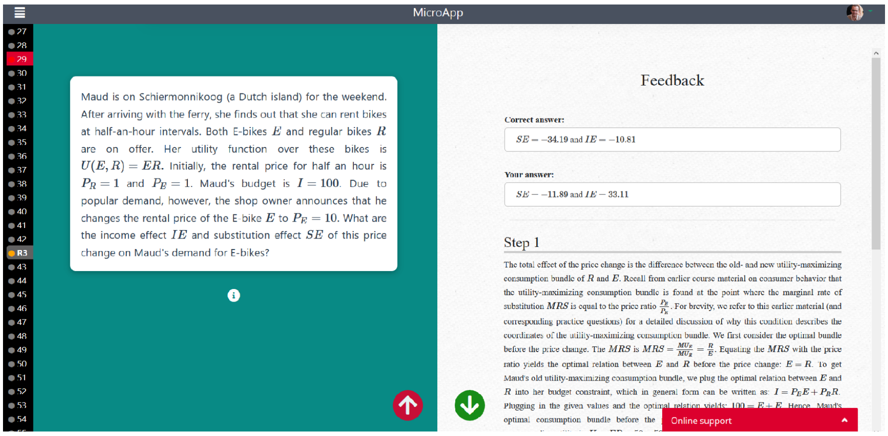
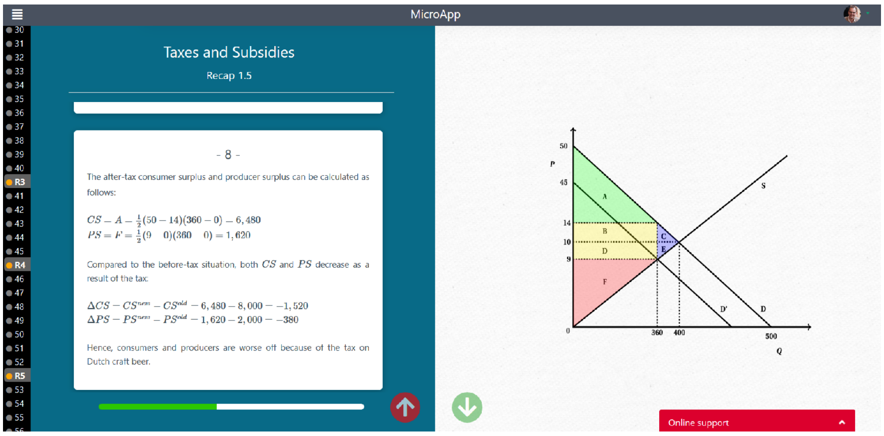


Figure A.3: Recap on the Course Topic of Taxes and Subsidies



A.2 A sample in-class survey (baseline period)

Figure A.4: Frontside of the Survey

Microeconomics for E&BE: Consumers & Firms
Weekly Survey

- Your answers to the survey questions remain private and will be handled confidentially.
- Your answers will be used solely for improving the course structure and for general research purposes.
- Please answer truthfully. Your answers will not in any way affect your final course grade.
- For questions about the survey, please contact the secretary: Kimberley Vudinh (k.m.vudinh@rug.nl).

This survey is about your study behavior in the past seven days (Saturday 15 September – Friday 21 September)

Q1: For this course, Microeconomics for E&BE, please indicate your approximate time investment in the past seven days for each of the following categories (where < means "less than"):

Category	Time investment
Attending classes (lecture, tutorial, practical)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Reading offline course material (e.g., textbook, lecture slides, hand-outs)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Practicing with offline course material (e.g., tutorial exercises, practice exams, etc.)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Use of the MicroApp	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Reading/watching online course material (not Microapp!) (e.g., YouTube tutorial videos)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Practicing with online course material (not Microapp!) (e.g., websites with practice exercises)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours
Any study activity that is not captured by one of the categories mentioned above (e.g., attending an office hour of a lecturer)	☐ Not at all ☐ Less than 1 hour ☐ 1 - <3 hour(s) ☐ 3 - <5 hours ☐ 5 hours or more than 5 hours

Figure A.5: Backside of the Survey

Q2: Please answer the following if you take other courses this block. With the categories from the previous question in mind, what was your approximate average time investment for these other courses (note again that < means "less than")?

- ☐ Not at all
- ☐ Less than 5 hours
- ☐ 5 - <10 hours
- ☐ 10 - <15 hours
- ☐ 15 hours or more than 15 hours

Example: let's say you follow two other courses: course A and course B. If your time investment last week was X hours for course A and Y hours for course B, your average time investment for these courses was: $\frac{X+Y}{2}$.

Thank you for filling in the survey!

- End of the survey -

A.3 Experimental variation in the in-class surveys

Figure A.6: Experimental Variation in the In-Class Surveys

Your MicroApp study goal [All experimental conditions]

Q3: Challenge yourself by setting a goal for how many hours you intend to make use of the MicroApp during your self-study in the upcoming week. Fill in the blanks below. Note that half hours can be written as 0.5.

In the upcoming seven days, my goal is to practice for ……… hours on the MicroApp before Friday 19 October 23:59.

- From tomorrow on, you can track your progress toward your study goal via www.microapp.nl.

Your preferred MicroApp study schedule [Conditions II and III]

Q4: In order to reach your MicroApp study goal, please indicate for each day in the table below your daily goal in hours, and how you intend to allocate these hours over the day. Make sure that your daily goals add up to your study goal stated in Q3. Note again that half hours can be written as 0.5.

	Saturday 13 October	**Sunday 14 October**	**Monday 15 October**	**Tuesday 16 October**	**Wednesday 17 October**	**Thursday 18 October**	**Friday 19 October**
Daily goal in hours							
Night 00:00-06:00							
Morning 06:00-12:00							
Afternoon 12:00-18:00							
Evening 18:00-00:00							

- From tomorrow on, you can consult your preferred study schedule at any time via www.microapp.nl. [Conditions II and III]
- The MicroApp now contains a *pacesetter* feature. The pacesetter is a moving point that implements your study goal by moving exactly according to your own preferred study schedule. From tomorrow on, you can see at any time where you stand relative to your own pacesetter via www.microapp.nl. [Condition III]

A.4 Questions on course goals in the first in-class survey

Figure A.7: Survey Questions on Course Goals

Q3: Please answer the following sub-questions related to your goals in this course (Microeconomics for E&BE). Note that the Dutch grading system is based on a ten-point scale, where 1 means very poor and 10 means excellent.

Q3.1: Which overall grade are you aiming for in this course?

Q3.2: Which grade are you aiming for on the mid-term exam (Tuesday 2 October, 2018)?

Q3.3: Which grade are you aiming for on the final exam (Thursday 1 November, 2018)?

Q3.4: What are the chances of you participating in the final exam [0-100%]?

A.5 Intertemporal choices about real effort in the third in-class survey

Figure A.8: Intertemporal Choice Task About Real Effort

A research team at this faculty is currently recruiting students for completing an online survey. Students who complete the survey have **a chance of winning one of the 10 available €50 Bol.com gift certificates**. Bol.com (www.bol.com) is the leading web shop in the Netherlands for books, toys and electronics. The available gift certificates are **exclusively for students in this course**!

Please indicate your preferences for filling in the survey by **making your decision in <u>each</u> of the 10 choices below**. One of your choices may be the survey you will have to answer to have a chance of winning a €50 Bol.com gift certificate. Indicating your preferences below does not mean that you are required to complete the online survey (send to you via email). The date of receiving the gift certificate does not depend on when you finish the survey (as long as you completed before the deadline).

The survey can either be answered in the **next 5 days (short survey)** or in the **next 35 days (long survey)**

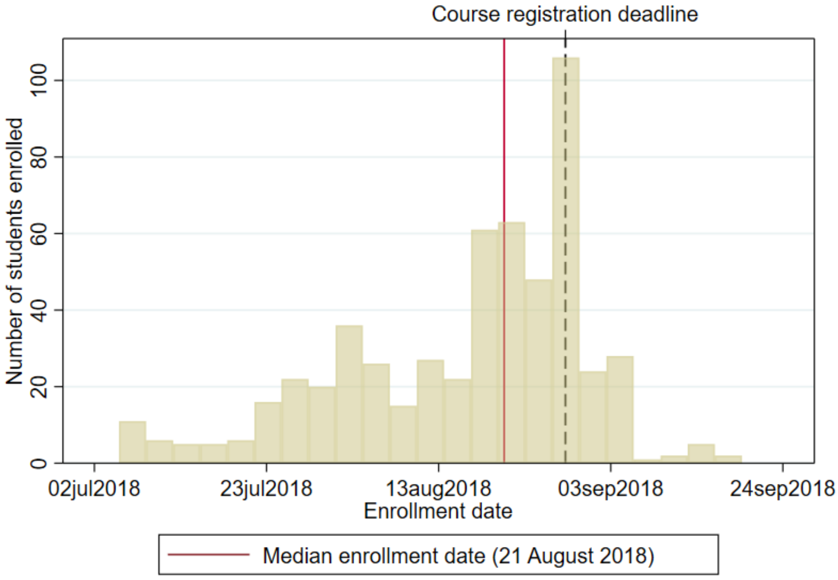
	Short version of the online survey **Before Thursday, 4 October 2018** *Length of the survey*		**Long version of the online survey** **Before Saturday, 3 November 2018** *Length of the survey*
Choice #1	○ 15 minutes	or	○ 30 minutes
Choice #2	○ 18 minutes	or	○ 30 minutes
Choice #3	○ 21 minutes	or	○ 30 minutes
Choice #4	○ 24 minutes	or	○ 30 minutes
Choice #5	○ 27 minutes	or	○ 30 minutes

The survey can either be answered in the **next 90 days (short survey)** or in the **next 120 days (long survey)**

	Short version of the online survey **Before Friday, 28 December 2018** *Length of the survey*		**Long version of the online survey** **Before Sunday, 27 January 2019** *Length of the survey*
Choice #6	○ 15 minutes	or	○ 30 minutes
Choice #7	○ 18 minutes	or	○ 30 minutes
Choice #8	○ 21 minutes	or	○ 30 minutes
Choice #9	○ 24 minutes	or	○ 30 minutes
Choice #10	○ 27 minutes	or	○ 30 minutes

A.6 Student enrollment dates in the course

Figure A.9: Student Enrollment Dates in the Course



Notes: The figure shows the student enrollment dates in the course. Students who enrolled before the registration deadline (26 August 2018) are guaranteed a spot in the course. Enrolling after the deadline is also possible, but it is not guaranteed that a spot is available. In practice, this policy is not too strict and students can generally still take the course when they enroll after the deadline.

A.7 MicroApp study time: intention-to-treat (ITT) estimates

Table A.1: Treatment Effects on MicroApp Study Time (Intention-to-Treat Estimates)

	(1)	(2)	(3)	(4)	(5)
Post=1	51.279***	51.279***	37.615***	37.615***	0.678***
	(9.549)	(9.553)	(6.181)	(6.183)	(0.151)
Post=1 × Treatment=2	-24.255*	-24.255*	-5.729	-5.729	-0.435*
	(13.663)	(13.668)	(9.029)	(9.030)	(0.239)
Post=1 × Treatment=3	-18.099	-18.099	-16.258*	-16.258*	-0.364
	(13.521)	(13.526)	(8.697)	(8.699)	(0.238)
Treatment=2		-3.268		-4.141	
		(13.630)		(8.685)	
Treatment=3		3.003		2.791	
		(13.357)		(8.831)	
Constant	90.588***	90.605***	52.208***	52.642***	4.265***
	(2.780)	(9.645)	(1.817)	(6.291)	(0.049)
R^2	.00866	.0088	.00758	.00783	.025
Number of students	309	309	569	569	305
Number of observations	2,472	2,472	4,552	4,552	1,174
Student fixed effects	Yes	No	Yes	No	Yes
Student inclusion/exclusion rules applied?	Yes	Yes	No	No	Yes
Time unit	Week	Week	Week	Week	Week
Dependent variable	Level	Level	Level	Level	Log

Notes: Identification of the intention-to-treat treatment effects on MicroApp study time (in minutes). Robust standard errors (clustered at the student level) are in parentheses. The variable *post* captures the experimental period. Study time during the transition day (5 October 2018) from the baseline period (lecture weeks 1-4) to the experimental period (lecture weeks 5-8) is excluded. There are two exclusion rules. First, a student is excluded when he or she participated in less than two practicals after the midterm exam. Second, a student is also excluded when he or she never made use of the MicroApp in the period until the first practical after the midterm exam. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

A.8 MicroApp study time: males versus females

Table A.2: Treatment Effects on MicroApp Study Time, by Gender

	(1)	(2)	(3)	(4)	(5)
Post=1	123.365***	138.784***	112.413***	130.489***	0.749***
	(32.740)	(31.150)	(30.764)	(28.530)	(0.203)
Post=1 × Treatment=2	-82.904**	-92.817**	-65.418*	-75.307**	-0.586*
	(41.698)	(39.346)	(39.514)	(36.888)	(0.322)
Post=1 × Treatment=3	-72.218*	-69.889*	-55.656	-56.767	-0.342
	(41.326)	(38.876)	(39.742)	(36.328)	(0.267)
Post=1 × Gender=1	-59.329	-53.658	-40.256	-58.137	-0.304
	(44.293)	(40.557)	(42.772)	(38.607)	(0.303)
Post=1 × Treatment=2 × Gender=1	71.775	74.742	50.428	91.183*	0.568
	(60.823)	(53.892)	(58.106)	(51.742)	(0.445)
Post=1 × Treatment=3 × Gender=1	102.298	144.450**	86.432	116.109*	-0.056
	(75.706)	(73.638)	(73.276)	(68.731)	(0.687)
Treatment=2		12.423		17.396	
		(30.014)		(27.421)	
Treatment=3		43.103		36.302	
		(29.115)		(26.518)	
Gender=1		18.225		11.925	
		(31.965)		(29.802)	
Treatment=2 × Gender=1		-40.660		-45.887	
		(45.304)		(42.203)	
Treatment=3 × Gender=1		-73.507		-51.724	
		(47.159)		(44.142)	
Constant	196.998***	174.118***	186.983***	165.859***	4.270***
	(6.303)	(20.401)	(6.355)	(18.525)	(0.049)
R^2	.0307	.0373	.0291	.0333	.0262
Number of students	300	300	404	404	300
Number of observations	1,148	1,148	1,343	1,343	1,148
Student fixed effects	Yes	No	Yes	No	Yes
Student inclusion/exclusion rules applied?	Yes	Yes	No	No	Yes
Time unit	Week	Week	Week	Week	Week
Dependent variable	Level	Level	Level	Level	Log

Notes: Identification of the gender-specific treatment effects on MicroApp study time (in minutes). The gender dummy variable is equal to one for female students, zero otherwise. Robust standard errors (clustered at the student level) are in parentheses. The variable *post* captures the experimental period. Study time during the transition day (5 October 2018) from the baseline period (lecture weeks 1-4) to the experimental period (lecture weeks 5-8) is excluded. There are two exclusion rules. First, a student is excluded when he or she participated in less than two practicals after the midterm exam. Second, a student is also excluded when he or she never made use of the MicroApp in the period until the first practical after the midterm exam. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

A.9 MicroApp study time: high-performance students versus low-performance students

Table A.3: Treatment Effects on MicroApp Study Time, by Pre-Experimental Performance

	(1)	(2)	(3)	(4)	(5)
Post=1	87.907***	102.566***	75.846***	81.445***	0.595***
	(27.085)	(24.748)	(26.052)	(23.421)	(0.186)
Post=1 × Treatment=2	-77.461**	-70.832**	-53.901	-39.209	-0.631**
	(36.758)	(33.434)	(35.800)	(32.043)	(0.257)
Post=1 × Treatment=3	-37.703	-13.136	-26.184	3.921	-0.540**
	(38.824)	(33.827)	(37.687)	(32.708)	(0.271)
Post=1 × Low performance=1	41.061	40.109	58.778	65.040	0.188
	(48.015)	(44.143)	(45.336)	(40.481)	(0.311)
Post=1 × Treatment=2 × Low performance=1	50.160	13.251	16.312	-11.579	0.546
	(62.607)	(58.321)	(59.175)	(53.961)	(0.514)
Post=1 × Treatment=3 × Low performance=1	-18.220	-46.720	-21.016	-70.685	0.453
	(67.866)	(64.721)	(65.449)	(59.222)	(0.501)
Treatment=2		3.946		4.358	
		(31.646)		(29.465)	
Treatment=3		2.903		-0.427	
		(29.889)		(28.111)	
Low performance=1		-40.877		-52.316*	
		(31.145)		(28.543)	
Treatment=2 × Low performance=1		-15.155		-2.633	
		(43.406)		(40.572)	
Treatment=3 × Low performance=1		45.913		52.178	
		(46.252)		(42.375)	
Constant	195.454***	198.094***	184.573***	192.723***	4.266***
	(6.234)	(19.212)	(6.359)	(18.191)	(0.048)
R^2	.0236	.0366	.0214	.0331	.0125
Number of students	305	305	409	409	305
Number of observations	1,174	1,174	1,372	1,372	1,174
Student fixed effects	Yes	No	Yes	No	Yes
Student inclusion/exclusion rules applied?	Yes	Yes	No	No	Yes
Time unit	Week	Week	Week	Week	Week
Dependent variable	Level	Level	Level	Level	Log

Notes: Identification of the treatment effects on MicroApp study time (in minutes), conditional on the pre-experimental performance of students in the course. Robust standard errors (clustered at the student level) are in parentheses. Pre-experimental course performance is measured as the within-treatment relative performance of a student on the midterm exam. The bottom (top) 50% is the low-performing (high-performing) group. Only students that satisfy the inclusion rules of this study are considered when calculating relative performances. The low-performance dummy variable is equal to one when a student is in the within-treatment bottom 50% in terms of pre-experimental performance, zero otherwise. The variable *post* captures the experimental period. Study time during the transition day (5 October 2018) from the baseline period (lecture weeks 1-4) to the experimental period (lecture weeks 5-8) is excluded. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

A.10 MicroApp study time: procrastinators versus nonprocrastinators

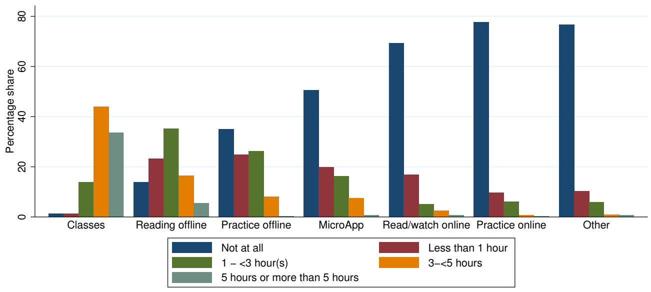
Table A.4: Treatment Effects on MicroApp Study Time, by Procrastination

	(1)	(2)	(3)	(4)	(5)
Post=1	94.569***	115.497***	94.206***	100.667***	0.651***
	(30.493)	(28.869)	(29.632)	(27.105)	(0.180)
Post=1 × Treatment=2	-67.784	-71.003*	-63.467	-42.742	-0.651**
	(41.933)	(38.544)	(40.293)	(37.015)	(0.306)
Post=1 × Treatment=3	-64.047	-53.109	-63.973	-45.828	-0.415
	(40.473)	(38.237)	(39.645)	(36.437)	(0.290)
Post=1 × Procrastinator=1	41.606	13.879	28.265	27.051	0.133
	(44.241)	(40.367)	(42.593)	(38.558)	(0.338)
Post=1 × Treatment=2 × Procrastinator=1	4.121	11.884	22.089	-1.653	0.472
	(59.140)	(53.566)	(56.669)	(51.379)	(0.498)
Post=1 × Treatment=3 × Procrastinator=1	38.448	50.474	62.671	45.302	0.082
	(65.930)	(60.896)	(63.961)	(57.237)	(0.514)
Treatment=2		7.797		3.849	
		(31.621)		(29.506)	
Treatment=3		32.417		32.080	
		(29.953)		(28.946)	
Procrastinator=1		-41.326		-49.517*	
		(29.011)		(26.424)	
Treatment=2 × Procrastinator=1		-23.501		-2.359	
		(41.780)		(38.822)	
Treatment=3 × Procrastinator=1		-26.660		-23.588	
		(44.233)		(40.380)	
Constant	195.932***	196.688***	184.252***	189.188***	4.266***
	(6.224)	(20.255)	(6.268)	(19.425)	(0.049)
R^2	.0212	.0408	.0188	.0363	.0185
Number of students	303	303	410	410	303
Number of observations	1,167	1,167	1,365	1,365	1,167
Student fixed effects	Yes	No	Yes	No	Yes
Student inclusion/exclusion rules applied?	Yes	Yes	No	No	Yes
Time unit	Week	Week	Week	Week	Week
Dependent variable	Level	Level	Level	Level	Log

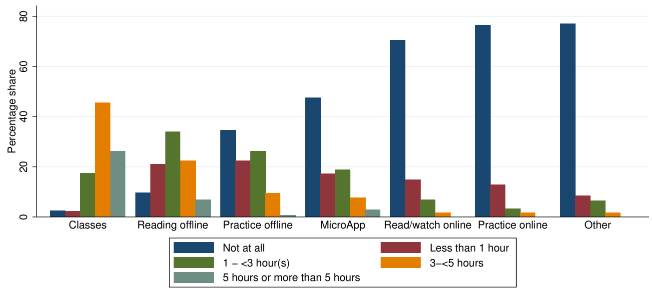
Notes: Identification of the treatment effects on MicroApp study time (in minutes), conditional on prior procrastination tendencies in the course. A student is identified as a procrastinator when his or her course enrollment date is after the within-treatment median course enrollment date. Robust standard errors (clustered at the student level) are in parentheses. Only students that satisfy the inclusion rules of this study are considered when calculating relative performances. The procrastinator dummy variable is equal to one when a student enrolled in the course after the within-treatment median course enrollment date, zero otherwise. The variable *post* captures the experimental period. Study time during the transition day (5 October 2018) from the baseline period (lecture weeks 1-4) to the experimental period (lecture weeks 5-8) is excluded. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

A.11 Second-stage time investments in learning inputs

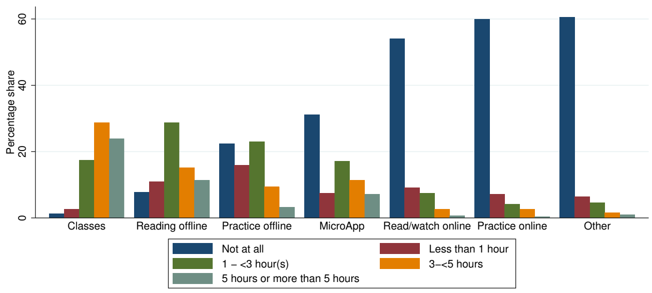
Figure A.10: Second-Stage Time Investments in Learning Inputs



(a) Lecture week 5



(b) Lecture week 6

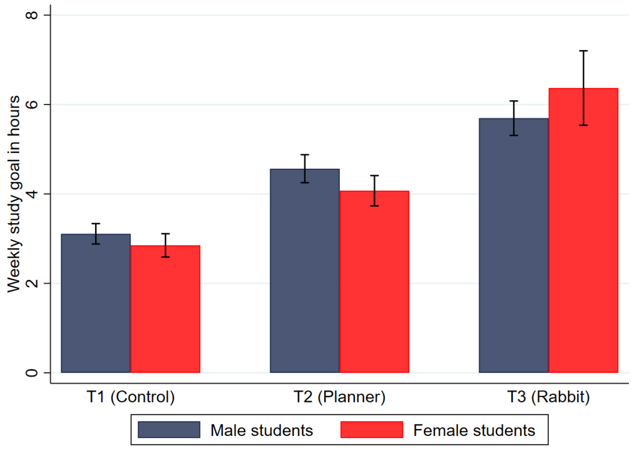


(c) Lecture week 7

Notes: Students' time investments in learning inputs are elicited at the end of each lecture week via in-class surveys. In the second stage of the experiment (overlapping with the last three lecture weeks and an exam week) there is experimental variation in the MicroApp and the in-class surveys (see Section 4.3). The final exam (worth 60% of the final grade) was in the exam week (the week after the seventh lecture week). No in-class survey was administered during the exam week. The learning input "classes" is partly incentivized (attendance is required to be eligible for a participation grade (worth 10% of the final grade)). The MicroApp is a newly introduced online learning platform in the course with abundant practice- and recap material. The categories "read/watch online" and "practice online" refer to non-MicroApp online study activities.

A.12 Goal-setting: males versus females

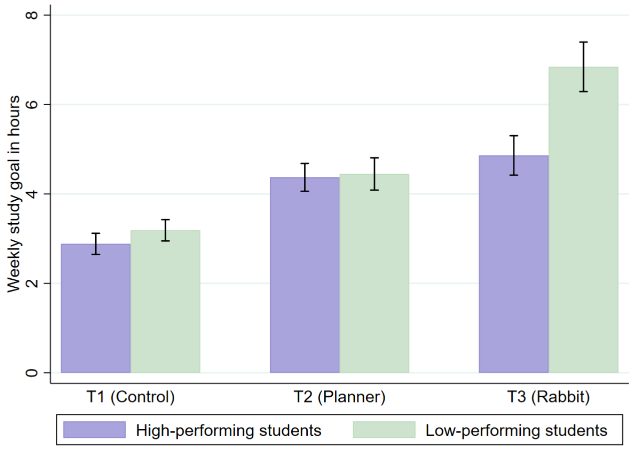
Figure A.11: Weekly Study Goals for MicroApp Usage, by Gender



Notes: The figure shows per experimental condition how many hours students on average plan to study on the MicroApp in each week during the experimental period, conditional on having set a non-zero study goal. Study goals are elicited through weekly in-class surveys. The average of these study goals is reported in the figure, split by gender and treatment status. The error bars denote ± 1 standard error.

A.13 Goal-setting: high-performance students versus low-performance students

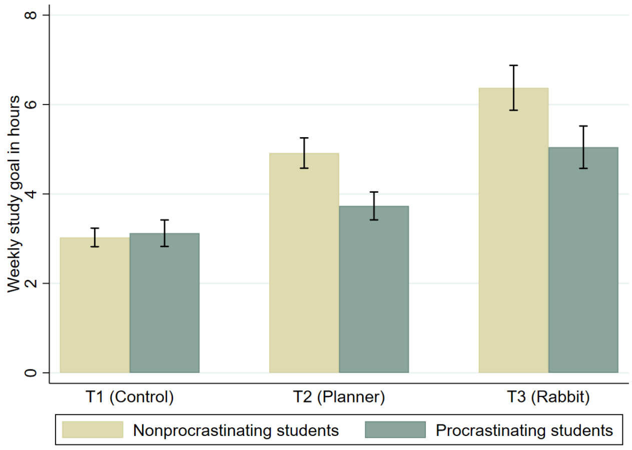
Figure A.12: Weekly Study Goals for MicroApp Usage, by Prior Performance



Notes: The figure shows per experimental condition how many hours students on average plan to study on the MicroApp in each week during the experimental period, conditional on having set a non-zero study goal. Study goals are elicited through weekly in-class surveys. The average of these study goals is reported in the figure, split by treatment status and students' pre-experimental performance in the course. Pre-experimental course performance is measured as the within-treatment relative performance of a student on the midterm exam. The bottom (top) 50% is the low-performing (high-performing) group. The error bars denote ± 1 standard error.

A.14 Goal-setting: procrastinators versus nonprocrastinators

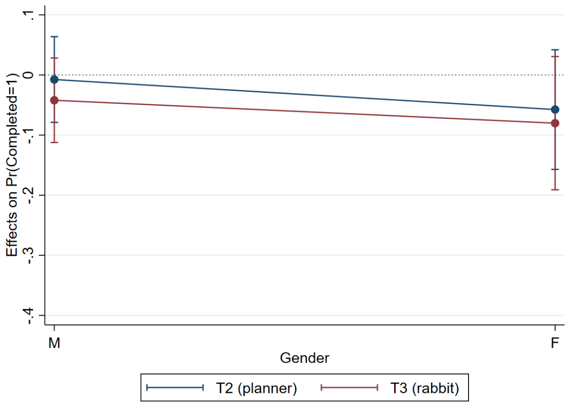
Figure A.13: Weekly Study Goals for MicroApp Usage, by Prior Procrastination



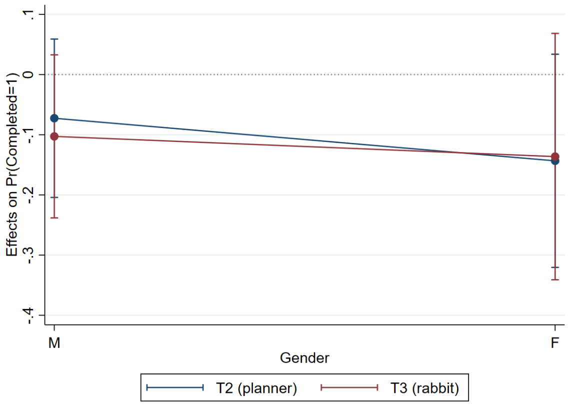
Notes: The figure shows per experimental condition how many hours students on average plan to study on the MicroApp in each week during the experimental period, conditional on having set a non-zero study goal. Study goals are elicited through weekly in-class surveys. The average of these study goals is reported in the figure, split by treatment status and students' prior procrastination tendencies in the course. The within-treatment median course enrollment date is used as the cutoff for identifying procrastinators (applied after) and nonprocrastinators (applied before) in each treatment. The error bars denote ± 1 standard error.

A.15 Goal completion: males versus females

Figure A.14: Treatment Effects on Study Goal Completion (logit MEs), by Gender



(a) Conditional on weekly study goal > 0

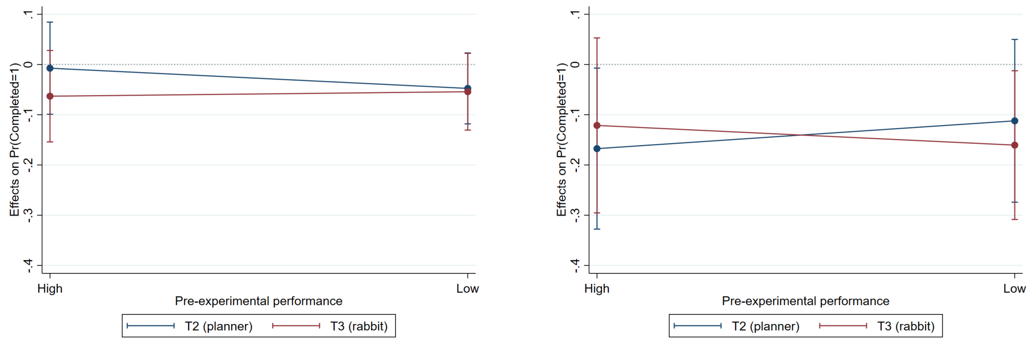


(b) Plus conditional on weekly MicroApp study time > 0

Notes: The figure shows the gender-specific treatment effects on the likelihood of achieving the study goal, after controlling for the goal level. The time period under consideration is the experimental period (lecture week 5 - exam week). The dependent variable is an indicator variable equal to one if a student completed his or her study goal for the week, zero otherwise. The coefficients shown are logit marginal effects. The reference group consists of students in the control group (T1). The vertical spikes indicate 95% confidence intervals.

A.16 Goal completion: high-performance students versus low-performance students

Figure A.15: Treatment Effects on Study Goal Completion (logit MEs), by Prior Performance



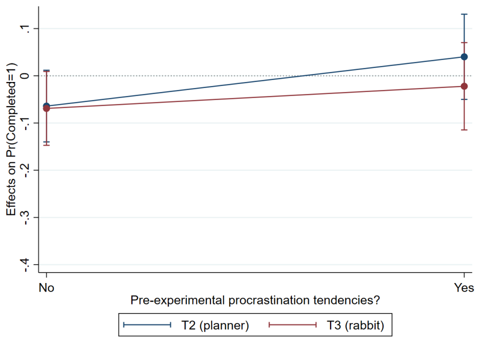
(a) Conditional on weekly study goal > 0

(b) Plus conditional on weekly MicroApp study time > 0

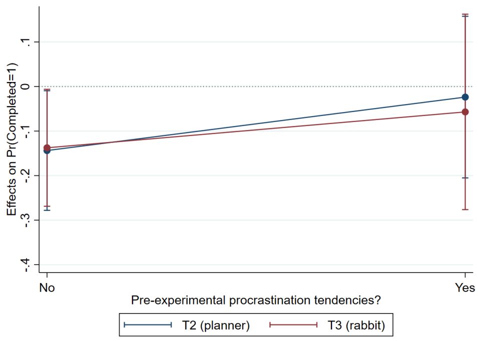
Notes: The figure shows the treatment effects on the likelihood of achieving the study goal, after controlling for the goal level. Treatment effects are shown separately for students with high/low pre-experimental performance. Low (high) performers are students in the within-treatment bottom (top) 50% based on performance on the midterm exam. The time period under consideration is the experimental period (lecture week 5 - exam week). The dependent variable is an indicator variable equal to one if a student completed his or her study goal for the week, zero otherwise. The coefficients shown are logit marginal effects. The reference group consists of students in the control group (T1). The vertical spikes indicate 95% confidence intervals.

A.17 Goal completion: procrastinators versus nonprocrastinators

Figure A.16: Treatment Effects on Study Goal Completion (logit MEs), by Procrastination



(a) Conditional on weekly study goal > 0



(b) Plus conditional on weekly MicroApp study time > 0

Notes: The figure shows the treatment effects on the likelihood of achieving the study goal, after controlling for the goal level. Treatment effects are shown separately for students with and without prior procrastination tendencies. I identify procrastination tendencies based on whether students enrolled for the course before or after the within-treatment median course enrollment date. The time period under consideration is the experimental period (lecture week 5 - exam week). The dependent variable is an indicator variable equal to one if a student completed his or her study goal for the week, zero otherwise. The coefficients shown are logit marginal effects. The reference group consists of students in the control group (T1). The vertical spikes indicate 95% confidence intervals.

A.18 Treatment effects on final exam performance: males versus females

Table A.5: Treatment Effects on Learning Outcomes (Final Exam Performance), by Gender

	(1)	(2)	(3)	(4)
Treatment=2	-0.457	-0.411	-1.250	-1.217
	(2.887)	(2.885)	(2.184)	(2.186)
Treatment=3	-0.341	-0.332	-2.970	-2.961
	(2.742)	(2.762)	(2.139)	(2.151)
Gender=1	7.035**	6.983**	3.470	3.368
	(3.477)	(3.486)	(2.988)	(2.976)
Treatment=2 × Gender=1	0.404	0.550	1.714	1.821
	(5.038)	(5.041)	(4.114)	(4.088)
Treatment=3 × Gender=1	-9.067*	-8.853*	-4.170	-3.950
	(5.100)	(5.137)	(4.473)	(4.473)
Constant	54.935***	54.243***	54.167***	53.143***
	(1.895)	(2.058)	(1.497)	(1.567)
Number of students	297	297	491	491
Student inclusion/exclusion rules applied?	Yes	Yes	No	No
Control for second-stage effort on the MicroApp?	No	Yes	No	Yes

Notes: The table reports the gender-specific treatment effects on learning outcomes. The dependent variable is the number of points scored on the final exam (out of 100 points). The indicator variable *gender* equals one for female students, zero otherwise. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

A.19 Treatment effects on final exam performance: high-performance students versus low-performance students

Table A.6: Treatment Effects on Learning Outcomes (Final Exam Performance), by Prior Performance

	(1)	(2)	(3)	(4)
Treatment=2	1.296	1.297	1.083	1.095
	(2.418)	(2.417)	(1.983)	(1.971)
Treatment=3	-3.765	-3.728	-4.412**	-4.436**
	(2.461)	(2.468)	(2.118)	(2.118)
Low performance=1	-18.581***	-18.577***	-18.244***	-18.171***
	(2.617)	(2.626)	(2.144)	(2.148)
Treatment=2 × Low performance=1	-4.415	-4.313	-4.625	-4.572
	(3.851)	(3.848)	(2.983)	(2.969)
Treatment=3 × Low performance=1	-0.804	-0.821	0.186	0.285
	(3.893)	(3.901)	(3.138)	(3.144)
Constant	65.073***	64.678***	63.898***	63.085***
	(1.754)	(1.943)	(1.450)	(1.529)
Number of students	302	302	508	508
Student inclusion/exclusion rules applied?	Yes	Yes	No	No
Control for second-stage effort on the MicroApp?	No	Yes	No	Yes

Notes: The table reports the treatment effects on learning outcomes. The dependent variable is the number of points scored on the final exam (out of 100 points). Treatment effects are shown separately for students with high/low pre-experimental performance (as measured by the grade on the midterm exam). Low (high) performers are students in the within-treatment bottom (top) 50% based on midterm scores. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

A.20 Treatment effects on final exam performance: procrastinators versus nonprocrastinators

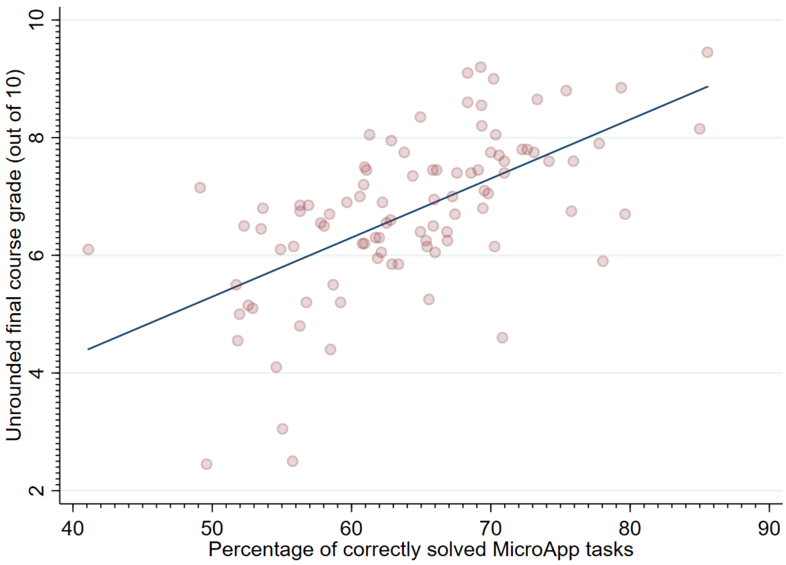
Table A.7: Treatment Effects on Learning Outcomes (Final Exam Performance), by Procrastination

	(1)	(2)	(3)	(4)
Treatment=2	0.253	0.373	-0.181	0.033
	(3.117)	(3.114)	(2.648)	(2.637)
Treatment=3	-2.742	-2.726	-2.242	-2.137
	(2.926)	(2.929)	(2.488)	(2.491)
Procrastinator=1	-2.542	-2.381	-3.234	-2.673
	(3.242)	(3.311)	(2.516)	(2.567)
Treatment=2 × Procrastinator=1	-2.279	-2.411	-1.476	-1.857
	(4.825)	(4.844)	(3.645)	(3.652)
Treatment=3 × Procrastinator=1	-0.692	-0.634	-3.718	-3.911
	(4.697)	(4.725)	(3.653)	(3.664)
Constant	58.179***	57.488***	56.567***	55.399***
	(2.042)	(2.358)	(1.843)	(1.992)
Number of students	301	301	508	508
Student inclusion/exclusion rules applied?	Yes	Yes	No	No
Control for second-stage effort on the MicroApp?	No	Yes	No	Yes

Notes: The table reports the treatment effects on learning outcomes. The dependent variable is the number of points scored on the final exam (out of 100 points). Treatment effects are shown separately for students with and without prior procrastination tendencies. I identify procrastination tendencies based on whether students enrolled for the course before or after the within-treatment median course enrollment date. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are in parentheses.

A.21 Correlation between student performance in the course and on the MicroApp

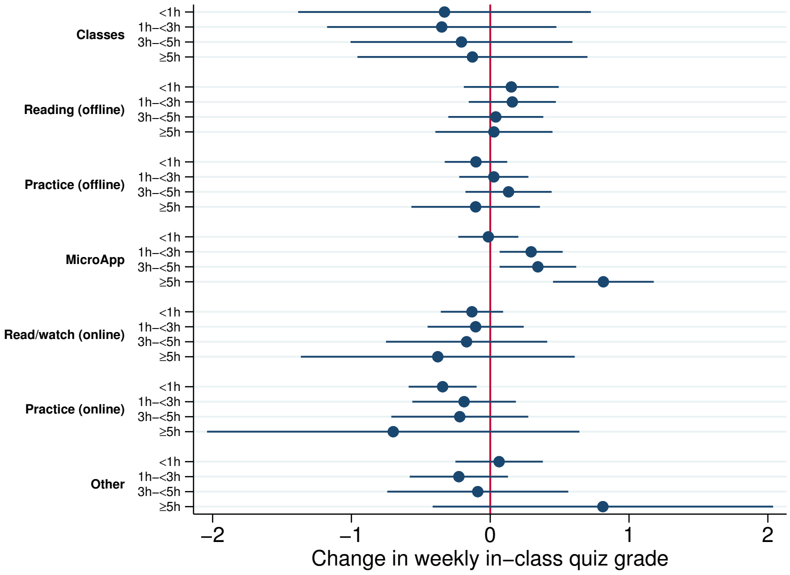
Figure A.17: Student Performance in the Course and on the MicroApp



Notes: The figure shows a scatterplot of student performance in the course (unrounded final course grade) and on the MicroApp (percentage of correctly solved practice tasks). The course grade is based on performance on a midterm exam (30%), a final exam (60%), and weekly in-class quizzes (10%). A higher grade indicates better performance. Only students who have solved at least 100 MicroApp practice tasks are included in the figure.

A.22 Time investments in learning inputs and performance on weekly quizzes

Figure A.18: Performance on Weekly Quizzes, by Time Investments in Learning Inputs



Notes: The figure plots the coefficients of a regression where the weekly in-class quiz grade (from 0 (poor) to 10 (excellent)) of a student is regressed on student fixed effects and on each level of categorical variables that capture students' weekly time investment in each learning input. The horizontal spikes indicate 95% confidence intervals. Standard errors are clustered at the student level. The reference category for each learning input consists of students with no time investment in that learning input during a specific week. The time investments of each student are elicited through end-of-week in-class surveys. Quizzes are administered at the end of each lecture week (Friday) and cover the course material of that week.