Self-Esteem and Procrastination: Evidence from a Field Experiment

Andrea Burro

This version: September 28, 2026

Click here for the most recent version

Abstract

Commitment devices are a common solution for procrastination, yet evidence on their effectiveness is mixed. The theory of self-handicapping procrastination points to a different cause: people procrastinate to protect their self-image, because failure after low effort can be attributed to effort rather than ability. This tendency is stronger for people with low self-esteem, so raising self-esteem should reduce procrastination. I test this mechanism in a field experiment with 2,774 university students. Students are randomly assigned to a self-esteem booster that provides a positive message about their study time, a goal-setting intervention, an information treatment that reports the same study time as the booster without the positive message, or a control group. All treatments are delivered through weekly notifications in a learning app. Using a novel high-frequency dataset from the app, I find that the booster increases overall study effort during the academic year, and that it raises study effort on every day of the week. The goal-setting and information interventions have no significant effect, so the result comes from the positive message rather than from information on study time. The effect on study effort persists in the following semester, after the intervention ends, which suggests habit formation. Survey evidence supports the mechanism: students in the booster group report higher self-esteem and lower procrastination.

Department of Economics, University of St. Gallen. andrea.burro@unisg.ch.

I am grateful to Beatrix Eugster for extensive discussions and her mentorship. I thank Palaash Bhargava, Luca Braghieri, Giovanni Burro, Bruno Caprettini, Michela Carlana, Matteo Fossi, Vincenzo Galasso, Sam Kang, John List, Paolo Piacquadio, Julia Seither, Sofia Shchukina, Sofía Sierra Vásquez, Giuseppe Sorrenti, Jeremia Stalder, and seminar participants at the University of St. Gallen, European Economic Review Summer School, Chicago School in Experimental Economics, Advances with Field Experiments Conference, North American Economic Science Association Conference, University of Chicago Booth Behavioral Economics Lab Meeting, University of Chicago Experimental Meeting, SKILS seminar, University of Chicago Behavioral Insights and Parenting Lab, LEAP lunch seminar, Milan PhD Economics Workshop, Swiss Society of Economics and Statistics Congress, European Economic Association Congress, European Association of Labour Economists Conference, European Economic Science Association Meeting, and Swiss Education Network Workshop for helpful comments and suggestions. I thank the University of Chicago and Bocconi University for hosting me while working on this project. The study was funded under the "Green Box programme", which is financed by the University of St. Gallen. The experiment was preregistered on the AEA RCT Registry on September 19, 2024 (RCT ID: AEARCTR-0014369) and approved by the Ethics Committee of the University of St. Gallen on May 31, 2024 (reference number: HSG-EC-20240508).

1 Introduction

In everyday life, people procrastinate on many tasks: writing a paper, going to the gym, paying taxes. Schraw, Wadkins, and Olafson (2007) define procrastination as the tendency to *"put off an impending task to a later time even when such practice results in counterproductive and needless delay."*

Procrastination is especially common among university students. Approximately 46% of university students identify as serious procrastinators (Solomon and Rothblum 1984); 80% to 95% have procrastinated at some point in their academic lives, and 50% regularly procrastinate on academic tasks (Steel 2007). Procrastinators are also more likely to have mental health issues and worse academic performance, although evidence mostly comes from correlational studies (De Paola and Scoppa 2015; Xu 2021; Arnold 2023; De Paola, Gioia, and Scoppa 2023). Understanding how to reduce procrastination therefore has direct implications for productivity and human capital accumulation.

Procrastination has traditionally been studied as a problem of self-control, in which people procrastinate because they undervalue the potential future benefits of completing an unpleasant task compared to the immediate costs of doing it. This has been modeled as quasi-hyperbolic discounting in β-δ models (e.g., Laibson 1997; O'Donoghue and Rabin 1999; Hyndman and Bisin 2022) or as a dual-self problem (Fudenberg and Levine 2006). These models suggest that commitment devices, such as binding deadlines and goal setting, should mitigate procrastination. However, empirical evidence on these interventions is mixed, which suggests that commitment devices alone may be insufficient to overcome self-control problems (Burger, Charness, and Lynham 2011; Bisin and Hyndman 2020; John 2020).

In psychology, procrastination is linked to self-esteem because it can be a self-handicapping strategy (Ferrari and Tice 2000). When individuals face tasks whose outcomes reflect on their ability, delaying effort can protect self-image: failure following low effort is attributed to a lack of effort rather than a lack of ability. This mechanism predicts that interventions that boost self-esteem should reduce self-handicapping procrastination.

In this paper, I test this prediction by estimating the causal effect of self-esteem on procrastination through a field experiment with two cohorts of first-year university students. I use a novel dataset from a learning app to measure procrastination and combine it with university administrative data and survey data. Students in the experiment are randomized into four groups: the *Booster* group, the *Goals* group, the *Information* group, and the *Control* group. The *Booster* group receives weekly notifications with a positive message about the student's study time. In the *Goals* group, the weekly notification asks students to set a goal for their study time. The *Information* group receives the same weekly summary of study time as the *Booster* group, without the positive message, and the *Control* group receives no additional notifications. Because the *Information* notification has the same timing and reports the same study time as the *Booster* notification, comparing the two arms separates the effect of boosting self-esteem from the effect of information on effort.

I find that boosting self-esteem increases effort in the app. The *Booster* treatment increases the number of correct answers, total answers, distinct questions answered, and time spent in the app by about 0.03 standard deviations. By contrast, the *Goals* and *Information* treatments have no significant effect on usage. The effect of the *Booster* is not concentrated on a specific day: point estimates are positive on every day of the week. One semester after the intervention, students in the *Booster* group still exhibit higher effort, which suggests

habit formation. Survey evidence supports the mechanism: students in the *Booster* group report higher self-esteem after the intervention and lower procrastination at the end of the academic year, while students in the other groups do not.

The main contribution of this paper is, to my knowledge, the first causal estimate of the effect of self-esteem on procrastination. Previous literature relied on correlational measures (e.g., Lay 1986; Tice and Baumeister 1997; Schuenemann et al. 2022), which may yield biased estimates due to reverse causality if higher procrastination reduces self-esteem. I address this by boosting self-esteem in a controlled experiment. This design relates to interventions by Hall, Zhao, and Shafir (2014), who find positive effects of self-esteem on cognitive performance, and Bursztyn et al. (2018), who find that increasing self-esteem reduces the demand for status goods.

I also contribute to the literature on self-control problems, which are prevalent across domains. DellaVigna and Malmendier (2006) examine gym members; Cadena et al. (2011) and Kaur, Kremer, and Mullainathan (2015) study workers; Thaler and Benartzi (2004) and Ashraf, Karlan, and Yin (2006) study savings; and Shah and Mukherjee (2026) retirement planning. In education, the evidence on commitment devices is mixed. Goal setting improves academic performance in some studies (Clark et al. 2020; Van Lent and Souverijn 2020), while similar interventions reduce it in others (Campos-Mercade, Thiemann, and Wengström 2025; Campos-Mercade and Wengström 2025). Deadlines are less effective: they have no effect on effort (Hyndman and Bisin 2022, 2026) or even reduce it (Burger, Charness, and Lynham 2011). However, committing to complete a task before the official deadline reduces last-minute behavior (Dirkmaat et al. 2023). More generally, the effects of goals and feedback depend on their design and on individuals' beliefs (Aulagnon et al. 2025; Abel et al. 2026; Brade,

Himmler, and Jäckle 2026; Strulov-Shlain and Wellsjo 2026). I add to this work by showing that boosting self-esteem increases effort in a setting where asking students to set goals does not. Comparing the *Booster* and *Information* arms also shows that the framing of feedback matters: the same information on study time has different effects depending on whether it comes with a positive message.

The granularity of the data allows me to estimate the evolution of the treatment effect over time. This matters for procrastination, which concerns the timing of effort. In a lab experiment, recording behaviors over long periods captures these dynamics (e.g., Augenblick and Rabin 2019; Fedyk 2024; Cerrone et al. 2026), but requires substantial resources and increases experimental fatigue. Using field data, Burger, Charness, and Lynham (2011) track study patterns over a semester but require monitored library logins, which could introduce demand effects. By contrast, participants in my setting likely perceive the notifications as a standard app update, which helps mitigate such effects. In addition, I observe repeated measures of study behaviors, while other field experiments took place in settings where individuals can procrastinate only once.

The data cover both the intervention and the following semester, after all students had switched back to the standard version of the app. Previous studies rarely estimate long-term effects; my setting allows me to test for habit formation, as Giuntella, Saccardo, and Sadoff (2026) do for the effect of sleep on academic performance.

My outcomes differ from those of previous studies. First, traditional outcomes such as grades or dropout would give only partial evidence, because effort shapes academic performance. My study is therefore similar to Cotton et al. (2026), Gneezy et al. (2019), and Ersoy (2023), which focus on students' effort rather than only on performance. Second, previous studies

have mainly measured beliefs. For example, training in emotion-focused strategies reduces perceived procrastination in a randomized experiment (Eckert et al. 2016). However, these results rely on validated procrastination scales, which may not accurately reflect actual behavior (Piedmont et al. 2000). With data on study effort in a high-stakes setting, I identify effects on both perceived and actual procrastination.

For educators, the results suggest that interventions providing periodic positive feedback or boosting self-esteem can reduce procrastination and raise academic productivity. These interventions do not require expensive ed-tech infrastructure; they can be implemented through standard teacher-student interactions. They are therefore a scalable tool for human capital accumulation across educational settings and grade levels (List 2024).

This paper proceeds as follows. Section 2 describes the setting and experimental design. Section 3 presents the main results. Section 4 examines the mechanism. Section 5 considers alternative explanations. Section 6 concludes.

2 Setting and experimental design

2.1 Setting

The university. The University of St. Gallen is a public Swiss institution offering bachelor's, master's, and doctoral programs in business, law, economics, and computer science. It enrolls roughly 10,000 students and is recognized as the leading business school in Switzerland and among the top 10 in Europe, according to the Financial Times (FT European Business School Rankings 2026).

Admission to the bachelor's program requires completion of the first year, known as the *assessment year*, which is offered in two tracks based on the language of instruction: German or English. All students take courses in business administration, economics, law, and mathematics. Each subject consists of a fall semester course and a spring semester course. To be admitted to the assessment year, students who do not have a Swiss high-school diploma have to take an entry test.

Grades range from 1 to 6, with 6 as the highest and 4 as the passing grade. Students may repeat the year once if they fail; failing a second time prevents them from enrolling in the same program anywhere in the country. These high stakes make the assessment year a well-suited context to investigate study effort.

Data partner. Brian AG is a Swiss company that provides educational services to universities and schools through its learning platform, Brian, where professors can upload study materials (e.g., flashcards, questions) that students access via smartphones and tablets.

The app uses gamification to increase student engagement with the topics (see Figure A.3 for an example of the interface). The interface resembles a journey, with different milestones representing each topic. To complete a milestone, students need to answer 20 questions. Students can answer as many questions as they want, including the same question multiple times.

Lecturers recommend Brian, but its use is voluntary and does not influence course grades. Introduced in 2020 in the Business Administration A course, the app is still used most extensively there, and students report it as helpful in course evaluations. Bieger et al. (2025) find that students who spend more time and answer more questions in Brian obtain higher exam grades, and that students who use the app more regularly during the semester perform

better. In this study, I analyze study behavior in Business Administration A and Economics A during the fall semester, and in Business Administration B during the spring semester.

2.2 Experimental design

Students were randomly assigned to four conditions: *Booster*, *Goals*, *Information*, and *Control*. The experiment covers two consecutive cohorts of first-year students, in the academic years 2024/2025 and 2025/2026. I used simple randomization to assign all eligible students in each cohort to the treatment arms; in the second cohort, the *Information* arm replaced the *Goals* arm.

Booster. Students in the self-esteem booster group received the standard version of the app until the start of the experiment (the last Sunday of the midterm break). After that, they received a positive notification about their app usage every Sunday at 1 p.m. for 11 weeks (up to the last Sunday before the Business Administration A exam). The message read: *"Great job! You studied for X minutes last week for Business Administration. Keep up the amazing work!"* (Panel A of Figure A.4). If students never opened the app during the week, they did not receive the notification.

Information. Every Sunday at 1 p.m. over the same 11 weeks, students in this arm received a notification reporting their study time in the previous week. The message had the same structure as the *Booster* message, without the positive opening and closing: *"Weekly summary. You studied for X minutes last week for Business Administration."* (Panel B of Figure A.4). As in the *Booster* group, students who never opened the app during the week did not receive the notification.

Goals. Every Sunday at 1 p.m. over the same 11 weeks, students in this arm received a notification prompting them to set a goal for their app use in the following week. The message read: *"Take control of your week. How many minutes do you plan to study on Brian next week in Business Administration?"* Clicking on the notification opened an input box in the app, where students could set their goal for the week (Panel C of Figure A.4).

Control. Students in the control group received the standard version of the app, without additional notifications. This group provides a baseline to identify the causal effect of the treatments.

The *Information* group isolates the self-esteem channel. Students in the *Booster* and *Information* groups received a notification at the same time, under the same conditions, and with the same information on their study time; only the *Booster* notification contained the positive message. Any difference between the two groups can therefore be attributed to the positive framing rather than to feedback on effort.

I use the *Goals* group to compare the *Booster* treatment with a traditional commitment device. If procrastination is only a self-control problem, the commitment device should reduce it more than positive feedback.

Figure A.5 presents the experimental timeline. The experiment took place during the academic years 2024/2025 and 2025/2026, covering both the fall and spring semesters. The fall semester began in mid-September and continued until Christmas, with a two-week break at the end of October. The corresponding exam period started in mid-January and ended in mid-February. The spring semester ran from mid-February to mid-May, with a two-week break at the end of March. The spring exam period began in mid-June and concluded in mid-July.

I complemented the analysis with three survey waves in each academic year. The baseline

survey was conducted from the end of September until mid-October, the midline survey from mid-December until mid-January, and the endline survey from the end of April until the end of May.

The intervention started on the last Sunday of the midterm break in the fall semester and lasted for 11 weeks, until the last week before the Business Administration A exam. During the spring semester there was no intervention, and all students returned to the standard version of the app. Short-run effects refer to the fall semester, during the intervention, and long-run effects to the spring semester, after the intervention.

2.3 Data and sample description

University. Data on nationality, gender, age, and language of study were obtained from the university's administrative records.

Brian. I used app usage data, including the number of questions answered, the number of correct answers, the time spent in the app, and the time and day when students answered each question. Outcome variables pool app activity across all the courses using the app: Business Administration A and Economics A in the fall semester, and Business Administration B in the spring semester.

Surveys. Participation in the surveys was voluntary. Students could complete the surveys by scanning a QR code shown in class during lecture time. Faculty members advertised the survey, and the link was also available on the university platform where students access study materials. To incentivize participation, in each wave, three students who completed the survey were randomly selected to win 100 CHF (around 120 USD) each. For the text of

the surveys, see Section A.5.

I collected the following information: self-reported study effort, overconfidence, and expected grade in Business Administration. I also used validated scales to measure self-esteem (Rosenberg's Self-Esteem Scale, Rosenberg 1965), procrastination (a short version of the General Procrastination Scale, Sirois, Yang, and van Eerde 2019), and time management (a short version of the Time Management Behaviors Scale, Peeters and Rutte 2005). Each scale produces an aggregate score by summing item responses, after reverse-coding negatively worded items, with higher values indicating higher levels of the underlying trait. The General Procrastination Scale includes items such as "In preparing for some deadlines, I often waste time by doing other things" or "I often find myself performing tasks that I had intended to do days before". Rosenberg's Self-Esteem Scale includes items like "I feel that I have a number of good qualities" or "On the whole, I am satisfied with myself". Finally, the Time Management Behaviors Scale contains items like "I feel in control of my time" or "I finish top-priority tasks before going on to less important ones".

Sample description. All students in the assessment year (N=3,600) were eligible. I excluded students who had not downloaded the app at the time of randomization or who had notifications disabled. The randomization sample consists of 2,774 students, or 77% of the eligible population: 900 in the *Control* group, 938 in *Booster*, 578 in *Goals*, and 358 in *Information*.

Table A.1 reports summary statistics for the experimental sample and the rest of the students enrolled in the assessment year. On average, students in the experiment are more likely to be enrolled in the German track (65% versus 51%) and to have Swiss nationality (75% versus 69%), while gender and age are balanced. As expected, students in the experiment used the

app more before the randomization. This reflects the inclusion criteria, since only students who had downloaded the app and enabled notifications were included in the experiment: 95% of the students in the experimental sample had used the app before the randomization, compared to 27% of the other students.

Table A.2 reports summary statistics for the experimental sample. Around three quarters of the students have Swiss nationality, and 65% are enrolled in the German track. Female students make up only 37% of the sample. In the baseline survey, the average expected grade in Business Administration is 4.9, with 6 being the highest grade and 4 the passing grade.

Table 1 shows no evidence of selection into treatments. For each baseline characteristic, I regress the variable on the treatment dummies and report the p-value of the F-test for joint significance. Gender, age, nationality, baseline app usage, and survey measures (self-esteem, procrastination, time management, and expected grade in Business Administration) are statistically balanced across groups. The only exception is the language of study ($p = 0.04$), which I include as a control in all regressions. Because the baseline survey was conducted before randomization and students were randomly assigned to the treatment arms, the groups are comparable at baseline.

Students could disable the notifications or uninstall the app. A small share of students did so during the experiment (Figure A.6). Until the second-to-last week of the experiment, fewer than 10% of the students in each treatment arm had disabled notifications or uninstalled the app. In the last week, the share rose to 10% in the *Booster* group and 15% in the *Information* group, while it remained below 1% in the *Goals* group.

The survey take-up rate is relatively low: 482 students in the experimental sample completed the baseline survey, 227 the midline survey, and 246 the endline survey. Table A.3 shows no

Table 1: Balance across treatment arms

Variable	Control	Booster	Information	Goals	p-value
Female	0.38	0.37	0.34	0.38	0.84
Age	19.77	19.78	19.67	19.82	0.99
Swiss	0.76	0.74	0.78	0.73	0.87
German Track	0.61	0.65	0.71	0.66	0.04
Brian User	0.95	0.95	0.95	0.93	0.54
Total Questions	783.57	780.71	657.11	766.96	0.40
Total Correct Questions	558.56	561.77	473.33	553.46	0.52
Total Unique Questions	412.92	419.96	368.18	412.49	0.59
Time Spent (minutes)	660.43	669.56	557.68	734.45	0.79
Procrastination	26.63	24.97	24.72	26.78	0.17
Self-Esteem	38.31	38.41	39.31	38.45	0.62
Time Management	33.70	35.09	35.41	34.54	0.23
Expected Grade (highest=6)	4.97	4.90	4.91	4.85	0.35
N (experimental sample)	900	938	358	578	
N (baseline survey)	147	182	68	85	

This table reports means for baseline variables for individuals in the experiment by treatment arm. The variables refer to the period before the randomization. Procrastination, self-esteem, time management, the expected grade in Business Administration A, refer to the baseline survey. Time spent is the total time in the Business Administration A and Economics apps before the first notification, in minutes, winsorized at the 99th percentile. The last column reports the p-value of an F-test of joint significance of the treatment arm indicators, controlling for experiment wave fixed effects, with heteroskedasticity-robust standard errors.

evidence that the composition of respondents differs across the three waves by demographics or baseline app usage. The only significant differences across waves are a higher proportion of German-track students in the baseline sample (70%) compared to the midline (60%) and endline (57%) samples, and a lower share of *Information* students in the endline sample. I therefore control for language of study in the analysis.

3 Results

To assess the effect of the treatment arms, I estimate the following regression:

$$Y_{it} = \alpha + \beta_1 \mathbf{1}\{\text{Booster}_i\} + \beta_2 \mathbf{1}\{\text{Goals}_i\} + \beta_3 \mathbf{1}\{\text{Information}_i\} + \gamma \mathbf{X}_i + \omega_t + \varepsilon_{it},$$

where Y_{it} denotes the outcome of student i in week t. Outcomes include (i) the number of correct answers, (ii) the total number of answers, (iii) the number of distinct questions answered, and (iv) the time spent in the app. Each outcome is standardized relative to the *Control* group, separately for each course and cohort. $\mathbf{X}_i$ is a vector of student-level controls measured at baseline, including gender, separate indicators for Swiss nationality and having a Swiss high-school diploma, language of study, and baseline usage of the app. The model includes week fixed effects (ω_t), and standard errors are clustered at the student level.

Given that students could disable the notifications, and that studying zero minutes in a week implies not receiving a notification in the *Booster* and *Goals* groups, the coefficients β_1, β_2, and β_3 capture the intention-to-treat (ITT) effects of assignment to the *Booster*, *Goals*, and *Information* interventions relative to the *Control* group.

Figure 1 shows that students assigned to the *Booster* group give more correct and total answers, answer more distinct questions, and spend more time in the app (Table A.4 reports the estimates). The estimated effects across these outcomes range from 0.029 to 0.036 standard deviations.[1] By contrast, there is no evidence that *Goals* or *Information* increase effort in the app.

1. Since I test the effect of each treatment on four outcomes, I adjust the p-values for multiple hypothesis testing. Using Romano-Wolf step-down p-values (Romano and Wolf 2005; see also List, Shaikh, and Xu 2019), the effect of *Booster* remains statistically significant at the 5% level for the number of correct answers and at the 10% level for the total number of answers, while the adjusted p-values for distinct questions and time spent are slightly above 0.1 (adjusted p-values in square brackets in Table A.4).

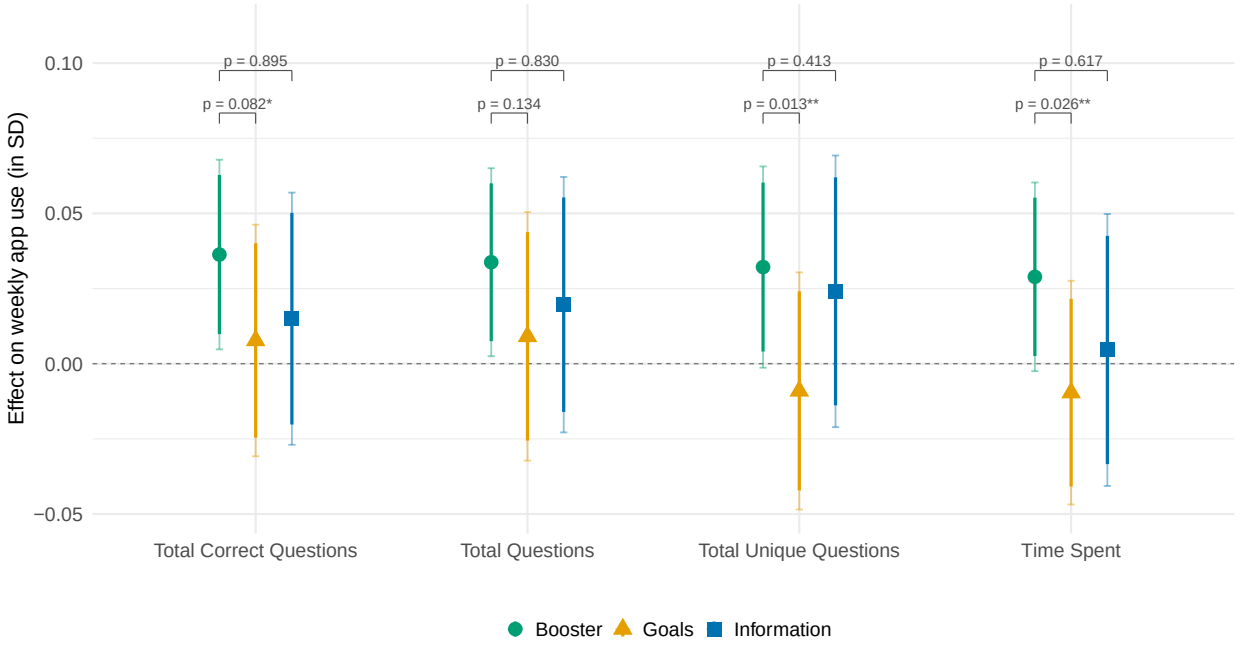


Figure 1: Effect of Booster, Goals, and Information on weekly app use. Each coefficient is the OLS estimate from the specification in Table A.4, with standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals. Brackets report the p-values of the tests of equality between the Booster coefficient and the Goals and Information coefficients. Outcome variables are standardized relative to the Control group.

The point estimates for *Information* are positive but smaller than those for *Booster*, and not statistically different from zero. The effect of *Booster* is larger than the effect of *Goals* for the number of correct answers ($p = 0.082$), the number of distinct questions ($p = 0.013$), and the time spent in the app ($p = 0.026$).

Higher effort need not mean less procrastination if students study more on one specific day only, for example right after receiving the notification. I therefore estimate the effect of the treatments on daily outcomes separately by day of the week. Figure 2 shows that the effect of the *Booster* on the number of answers is positive on every day of the week, and statistically significant on four of them (Tuesday, Friday, and Sunday at the 5% level; Saturday at the 10% level). The effect is not concentrated on Sundays, when the notification is sent, which suggests

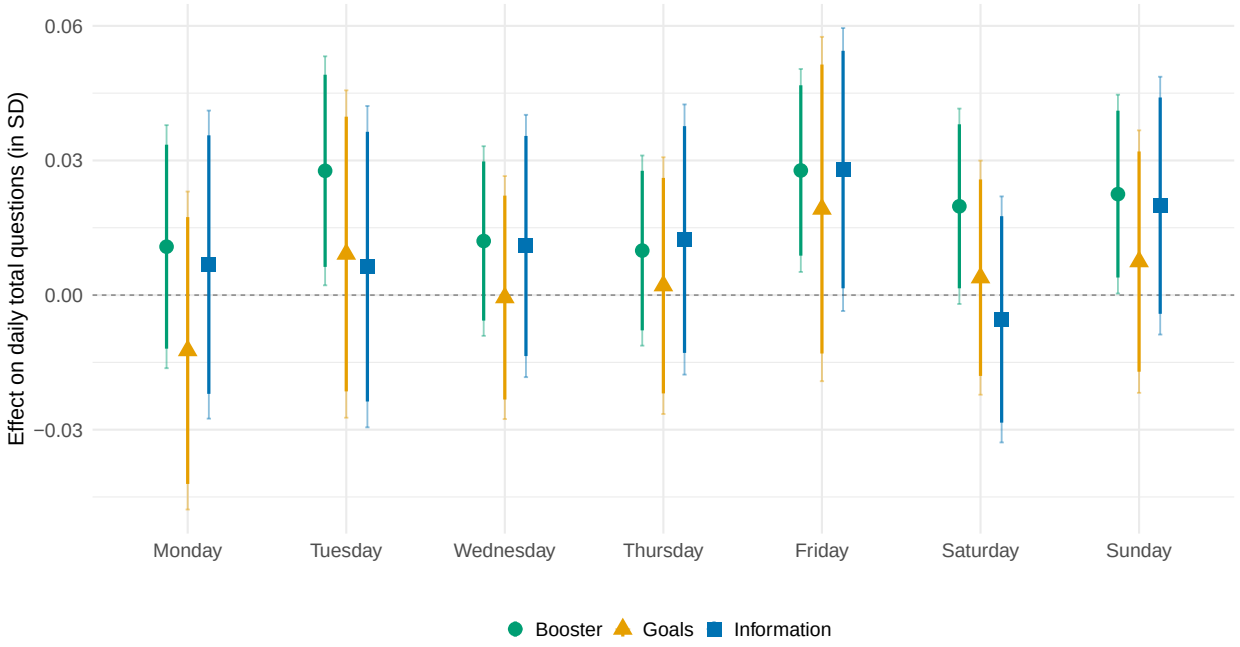


Figure 2: Effect of Booster, Goals, and Information on the daily number of answers, by day of the week. Each coefficient is the OLS estimate relative to the Control group, with standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals. The outcome variable is standardized.

that students in the *Booster* group distribute their effort across the week and procrastinate less. The estimates for both *Goals* and *Information* are mostly not statistically different from zero.

Figure 3 splits the effect of the *Booster* into the short run, when students are exposed to the treatment, and the long run, when all students have returned to the standard version of the app. The overall effect is driven by the long run. However, the fall estimate reflects only partial exposure. Notifications accumulate over the treatment weeks, so in the fall semester students had received only part of the treatment, while by the spring semester they had received all of it. The spring estimate suggests that students start studying more while they are exposed to the intervention and keep studying more after the notifications stop. I interpret this pattern as evidence of habit formation.

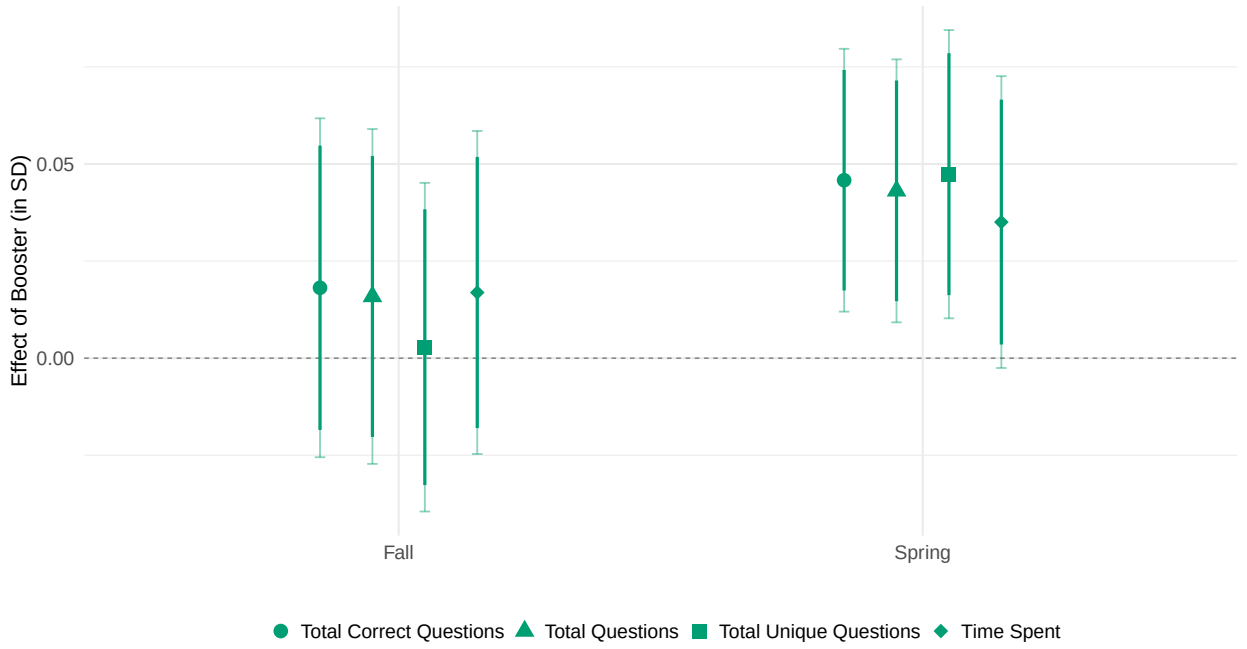


Figure 3: Effect of Booster in the short run (fall semester, during the intervention) and in the long run (spring semester, after the intervention). Each coefficient is the OLS estimate relative to the Control group, with week fixed effects, controls, and standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals. Outcome variables are standardized relative to the Control group.

The *Booster* changes both behaviors and beliefs. Figure 4 shows no difference across groups in self-reported procrastination in the midline survey, while students in the *Booster* group report lower levels of procrastination in the endline survey. The effect corresponds to a reduction of around 0.27 standard deviations and is statistically different from the effect of *Goals* ($p = 0.019$). Students in the *Goals* and *Information* groups do not differ from the *Control* group in reported procrastination.

If students procrastinate because they lack self-control, the *Booster* could also improve their time management. However, I find no significant effect on time management (Figure A.9). This suggests that students did not become better at managing their time, but study more in response to the positive feedback.

A possible side effect of sending positive feedback to all students is overconfidence. However, I find no statistically significant difference across treatment groups in the expected exam grade (Figure A.10).

Figure A.11 in the appendix reports the effects on each item of the procrastination scale, with reverse-coded items flipped so that higher values always mean more procrastination. In the endline survey, the point estimates for *Booster* are negative for all nine items, although close to zero for delaying before starting work and for buying essential items at the last minute. The largest reductions are for wasting time on other things before deadlines and for continually saying "I'll do it tomorrow", the two items that describe putting off work most directly. The endline effect on the scale is therefore not driven by a single item. In the midline survey, the estimates for *Booster* are negative for six items and positive for two reverse-coded items (finishing tasks sooner than necessary and accomplishing all that is planned for the day), in line with the null effect on the scale.

All estimates are ITT effects, and assignment to the *Booster* treatment does not mean that a student received all 11 *Booster* notifications: if a student never opened the app in a given week, no notification was sent. Moreover, the perceived impact of the message may vary with weekly effort. For example, receiving a message praising a student for studying only four minutes may not effectively boost self-esteem.

Figure A.7 displays, for each week, the share of students in the *Booster* group with zero app use, who therefore did not receive that week's notification, and the share with low but positive app use (under five minutes). The share with low but positive app use is small and stable over time (around 5%), whereas the share with zero app use is substantially higher and decreases over the semester, from 55% in the first week to 24% in the last week. If these

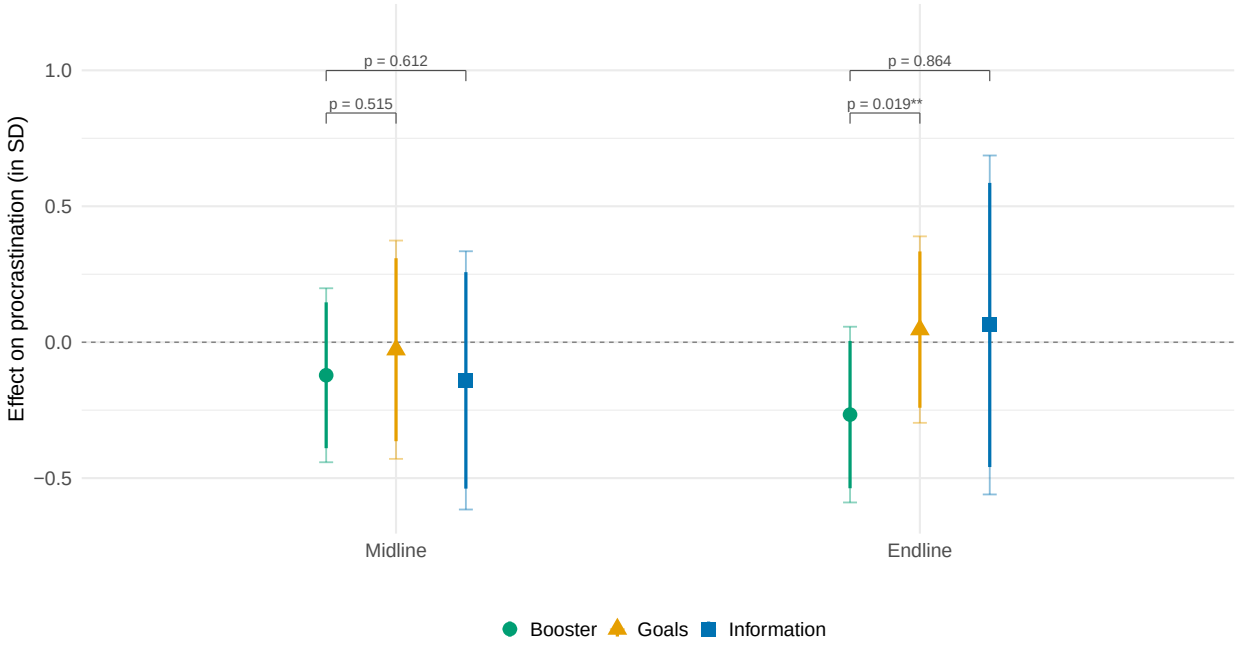


Figure 4: Effect of Booster, Goals, and Information on reported procrastination in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Brackets report the p-values of the tests of equality between the Booster coefficient and the Goals and Information coefficients. Reported procrastination is measured using the Short General Procrastination Scale (Sirois, Yang, and van Eerde 2019) and standardized. The full text of the questions is in Section A.5.

subgroups attenuate the treatment effect, the estimated ITT should be interpreted as a lower bound for both app-based study effort and self-reported survey outcomes.

Section A.1 reports the preregistered heterogeneity analyses. The effect of the *Booster* is positive for both male and female students, and similar for students with and without a Swiss high-school diploma.

4 Mechanism: self-esteem

In economics, procrastination is usually explained by present bias. In the quasi-hyperbolic β-δ model (Laibson 1997; O'Donoghue and Rabin 1999), a student gives extra weight to the

present, so the immediate cost of studying outweighs its discounted future benefits and the student delays. The present-bias parameter β is a fixed preference. A message that praises past study time without changing the costs or benefits of studying should therefore leave behavior unchanged, and the natural remedy for procrastination is a commitment device.

Self-handicapping offers a different account. Individuals derive utility from their beliefs about their own ability (Bénabou and Tirole 2002; Kőszegi 2006), and these beliefs can be fragile (Kőszegi, Loewenstein, and Murooka 2022). A student who believes that the outcome of an exam depends on both ability and effort learns that their ability is low if they fail after studying hard. If they fail after procrastinating, they can blame the lack of effort instead. Delaying effort therefore protects the student's self-image, and this protection is more valuable when self-esteem is low.

I capture this idea in reduced form by letting present bias depend on self-esteem. Section A.2 presents the model in full and relates it to the standard β-δ framework. The utility of a student at time t, evaluating a stream of payoffs $(u_t, u_{t+1}, \ldots, u_T)$, is

$$U_t(s) = u_t + \beta(s) \sum_{\tau=1}^{T-t} \delta^\tau u_{t+\tau},$$

where s is the student's current self-esteem, $0 < \delta \leq 1$ is the exponential discount factor, and $\beta(s) \in (0, 1]$ is a present-bias parameter that depends on self-esteem, with $\beta'(s) > 0$. A lower $\beta(s)$ captures the pull toward delay created by the need to protect one's self-image: a student with low self-esteem behaves as if more present-biased and is more likely to procrastinate, while as s increases, $\beta(s) \to 1$ and behavior approaches exponential discounting.

The framework predicts that the *Booster*, by raising s, increases $\beta(s)$ and reduces procrastina-

tion. The *Information* notification reports the same study time without the positive message, so it should not change s and should not reduce procrastination through this channel. I test the channel directly with the measures of self-esteem collected in the surveys.

Self-esteem was balanced across groups at baseline (Table 1). After the intervention, students in the *Booster* group reported significantly higher self-esteem in the midline survey (0.36 standard deviations), and the effect remains positive in the endline survey (0.29 standard deviations, statistically significant at the 10% level), as shown in Figure 5. By contrast, the *Information* treatment has no effect on self-esteem, and the difference between *Booster* and *Information* is statistically significant both in the midline ($p = 0.024$) and in the endline survey ($p = 0.032$). Since the two notifications differ only in the positive message, this comparison supports the interpretation that the *Booster* operates through self-esteem rather than through information on study time.

Figure A.12 in the appendix shows the effects on each of the ten items of the Rosenberg scale, again with reverse-coded items flipped. The point estimates for *Booster* are positive for every item in both surveys, so the effect on the scale reflects a broad shift in how students evaluate themselves rather than a change in one or two answers. In the midline survey, the largest estimates, around 0.35 standard deviations, are for doing things as well as most people, not feeling like a failure, and having something to be proud of. These items relate to competence and failure, the dimension of self-esteem that self-handicapping is meant to protect. The item-level estimates for *Information* are imprecise, and in the endline survey they are negative for all ten items.

Students in the *Goals* group do not differ from the *Control* group in reported self-esteem either.

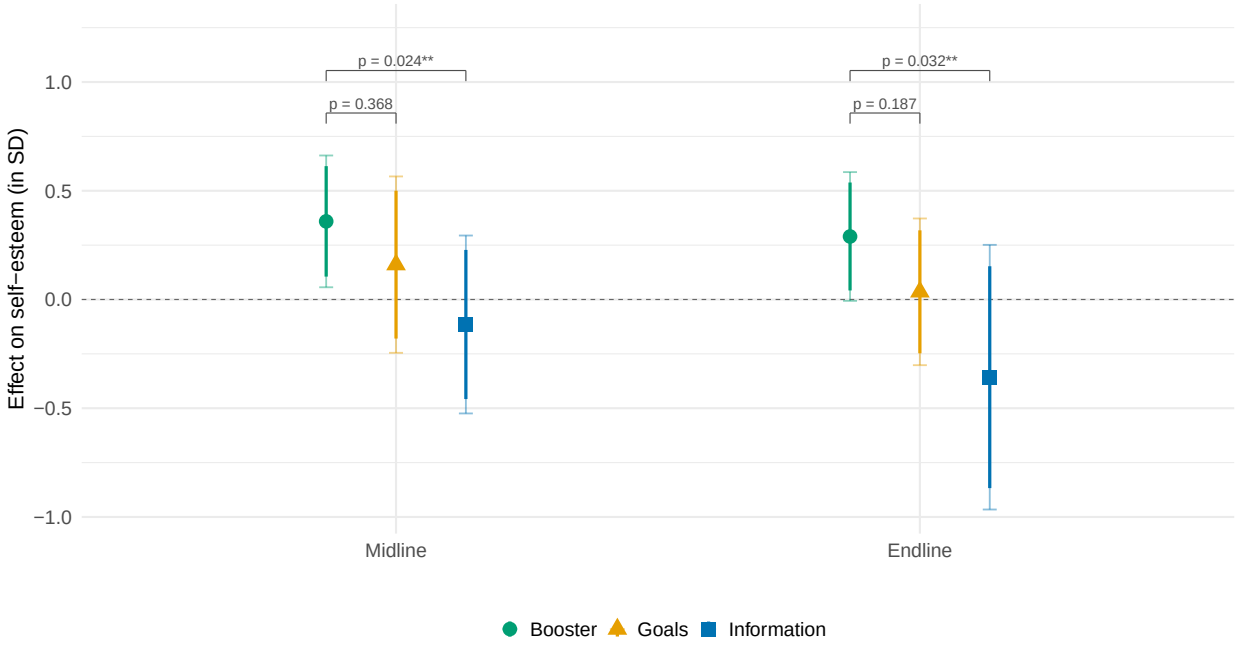


Figure 5: Effect of Booster, Goals, and Information on reported self-esteem in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Brackets report the p-values of the tests of equality between the Booster coefficient and the Goals and Information coefficients. Self-esteem is measured using the Rosenberg Self-Esteem Scale (Rosenberg 1965) and standardized. The full text of the questions is in Section A.5.

5 Alternative explanations

5.1 Spillovers

Spillovers could bias the ITT estimates if students study in groups and the treatment assignment of one student affects the effort of another student in the same group. For example, a student in the *Control* group who sees a peer in the *Booster* group being praised for their effort may question their own performance and study less.

To test for spillovers, I use randomly formed groups to which students were assigned during a mandatory orientation week. Thiemann (2022) shows that these groups matter for friendship networks, major choices, GPA, and dropout rates. The groups, with an average size of

24 students, were formed by randomization stratified by gender and by whether students completed high school in Switzerland. During the orientation week, each group spent 60 hours together and prepared a case-study presentation, which it delivered on the last day. Figure A.8 reports a regression of study effort on the share of peers in each treatment arm within the orientation week group, controlling for the student's own treatment assignment. None of the coefficients is statistically significant, so there is no evidence of spillovers.

5.2 Salience of the app and engagement

Another potential concern is that the effects are driven by greater salience of the app rather than by higher self-esteem. Students in the *Booster* group receive weekly notifications, which could increase exposure to the app relative to the *Control* group and mechanically boost engagement. However, students in the *Goals* and *Information* groups also receive weekly notifications at the same time. In particular, the *Information* notification is identical to the *Booster* notification except for the positive message, so any salience effect should similarly affect this group.

Since the *Goals* and *Information* treatments do not have the same effects as the *Booster*, the positive effect of the self-esteem booster is unlikely to be due only to increased app salience. Moreover, students in the *Booster* group received on average fewer notifications than those in *Goals*: by design, students in *Booster* received the notification only if they opened the app during the week, while those in *Goals* received it regardless of their effort. Despite receiving fewer notifications, the *Booster* group differs from the *Goals* group both in observed behavior (Figure 1) and in beliefs (Figure 5 and Figure 4).

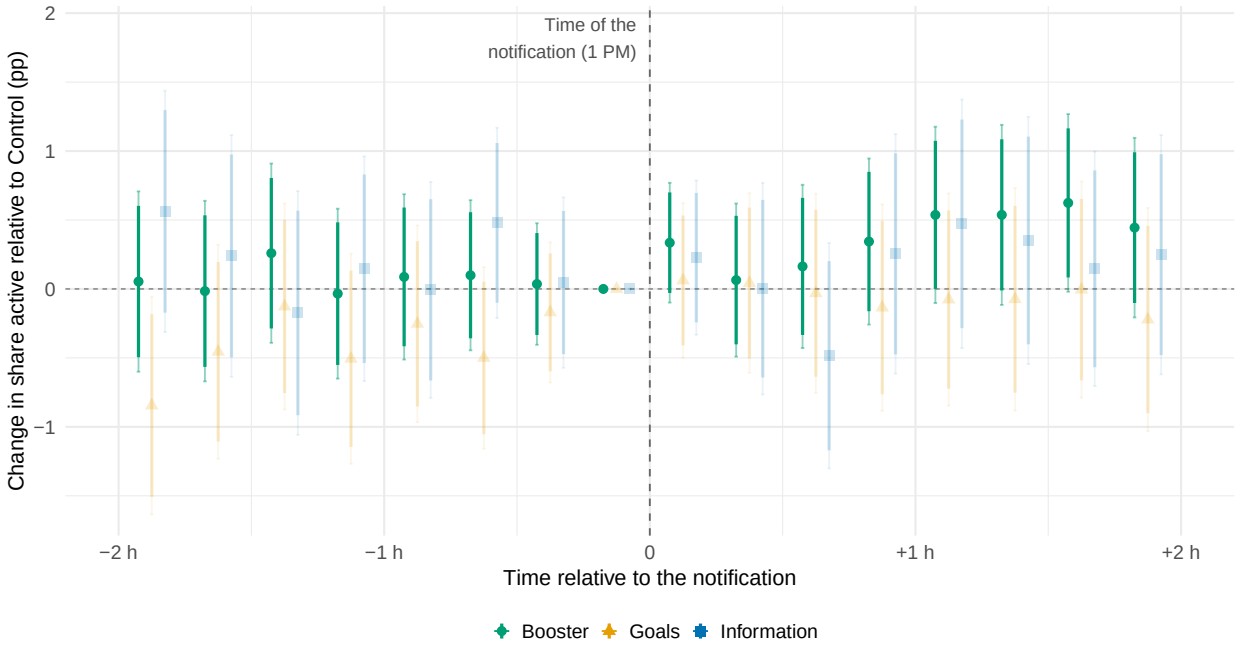


Figure 6: Change in the share of students active in the app in 15-minute intervals around the Sunday notification (1 p.m.), relative to the Control group and to the interval before the notification, during the weeks of the intervention. Thick and thin bars represent 90% and 95% confidence intervals.

To further explore this concern, Figure 6 plots the change in the share of students active in the app in 15-minute intervals around the notification on Sundays, relative to the *Control* group and to the interval before the notification. Before the notification, there is no difference between the treatment arms and the *Control* group. After the notification, the share of active students increases for students in the *Booster* group, by around 0.5 percentage points in the second hour, while it does not change for students in the *Goals* and *Information* groups. Since the *Information* notification is equally salient, these patterns indicate that the effect of the self-esteem booster is not simply a mechanical consequence of app reminders.

6 Conclusion

This paper tests whether an intervention that increases self-esteem can reduce procrastination. I conducted a field experiment with two cohorts of first-year university students, in collaboration with a learning app.

The results indicate that the self-esteem booster increases study effort in the app among both male and female students. Students in the *Booster* group also exhibit more effort in the semester after the intervention. However, students in the *Goals* and *Information* groups do not increase their effort. Since the *Information* notification reports the same study time as the *Booster* notification, the results suggest that the positive message, rather than feedback on effort or reminders to use the app, drives the effect.

For the design of educational interventions, the results suggest that the framing of feedback matters: the same information can have different effects depending on the message that accompanies it. They also suggest that low-cost interventions, such as periodic positive feedback, can have positive long-term effects on student effort.

One limitation of this study is that the analysis focuses on online learning, and future studies should also analyze offline learning. However, given the proliferation of learning platforms, I expect future generations to spend more time using new learning tools, which makes this context increasingly relevant. In an ideal setting, a researcher would have access to both online and offline learning data. In my setting, this could have been possible if library access data were available, assuming students primarily study there. Since students do not scan their student cards to access the library, it is not possible to objectively quantify offline effort in this study.

Regarding external validity, future studies should replicate these findings. While the theoretical framework suggests that these results should generalize, individuals in other settings might respond differently. First, students at the start of university have recently transitioned to a new learning environment; they are still adapting their learning strategies and may be more willing to change them (Dai, Milkman, and Riis 2014). Second, compared to other populations, university students may attach greater weight to academic performance in their self-image, given the strong implications of a degree for future earnings and social status.

Not all students assigned to the *Booster* group received full exposure to the treatment. Future interventions incorporating teacher-delivered positive messages could overcome this problem, as face-to-face interactions might achieve higher treatment compliance than in-app notifications.

Finally, the results show a causal effect of self-esteem on effort among students. A natural extension would be to boost self-esteem in other settings where individuals might self-handicap to protect their self-image, such as newly hired workers or freelancers working on a first project with a new client. A freelancer, for example, faces a trade-off: high effort increases the chance of future work, but low effort provides a self-protective excuse (blaming effort rather than ability) if the contract is not renewed.

References

Abel, Martin, Tomoko Harigaya, Michael Kremer, and Jessica Zhu. 2026. "The Effect of Reminders for Self-Set Goals on Productivity." *Journal of Development Economics* 182:103822.

Arnold, Ivo J. M. 2023. "The Link between Procrastination and Graduation Rates: Evidence from the ALEKS Learning Platform." *Education Economics* 31 (3): 275–287.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121 (2): 635–672.

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

Augenblick, Ned, and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies* 86 (3): 941–975.

Aulagnon, Raphaëlle, Julian Cristia, Santiago Cueto, and Ofer Malamud. 2025. "Streaks to Success: The Effects of Highlighting Streaks on Student Effort and Learning." *Economics of Education Review* 109:102721.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics* 117 (3): 871–915.

Bieger, Thomas, Samuel Heer, Simon Kuster, Dario Mitterer, and Yannik Breitenstein. 2025. *Impact of Gamified Learning on Learners' Performance: Time and Learning Time Patterns Matter.* Working Paper.

Bisin, Alberto, and Kyle Hyndman. 2020. "Present-Bias, Procrastination and Deadlines in a Field Experiment." *Games and Economic Behavior* 119:339–357.

Brade, Raphael, Oliver Himmler, and Robert Jäckle. 2026. "No Student Left behind? Relative Feedback and University Completion." *Journal of Economic Behavior & Organization* 241:107383.

Burger, Nicholas, Gary Charness, and John Lynham. 2011. "Field and Online Experiments on Self-Control." *Journal of Economic Behavior & Organization* 77 (3): 393–404.

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao. 2018. "Status Goods: Experimental Evidence from Platinum Credit Cards." *The Quarterly Journal of Economics* 133 (3): 1561–1595.

Cadena, Ximena, Antoinette Schoar, Alexandra Cristea, and Héber M. Delgado-Medrano. 2011. *Fighting Procrastination in the Workplace: An Experiment.* Technical report 16944. National Bureau of Economic Research.

Campos-Mercade, Pol, Petra Thiemann, and Erik Wengström. 2025. *Performance Incentives in Education: The Role of Goal Mismatch.* Technical report 2025:5.

Campos-Mercade, Pol, and Erik Wengström. 2025. *Threshold Incentives and Academic Performance.*

Cerrone, Claudia, Anujit Chakraborty, Hyok Jung Kim, and Leonhard K. Lades. 2026. "Estimating Present Bias and Sophistication over Effort and Money." *Experimental Economics,* 1–23.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush. 2020. "Using Goals to Motivate College Students: Theory and Evidence From Field Experiments." *The Review of Economics and Statistics* 102 (4): 648–663.

Cotton, Christopher S., Brent R. Hickman, John A. List, Joseph Price, and Sutanuka Roy. 2026. "Why Don't Struggling Students Do Their Homework? Disentangling Motivation and Study Productivity as Drivers of Human Capital Formation." *Journal of Political Economy* 134 (1): 86–149.

Dai, Hengchen, Katherine L. Milkman, and Jason Riis. 2014. "The Fresh Start Effect: Temporal Landmarks Motivate Aspirational Behavior." *Management Science* 60 (10): 2563–2582.

De Paola, Maria, Francesca Gioia, and Vincenzo Scoppa. 2023. "Online Teaching, Procrastination and Student Achievement." *Economics of Education Review* 94:102378.

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

Delavande, Adeline, Emilia Del Bono, Angus Holford, Sonkurt Sen, and Vedran Lesic. 2026. "Expectations about the Productivity of Effort and Academic Outcomes: Evidence from a Randomized Information Intervention." *Journal of Human Resources.*

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719.

Dirkmaat, Thomas, Kirsten I. M. Rohde, Evelien van de Veer, Bram van Dijk, and Xiao Yu. 2023. "Managing 'Last Moment Behavior': Non-binding Target Dates to Reduce the Spikes in Task Completion at Deadlines." *Journal of Public Economics* 219:104813.

Eckert, Marcus, David D Ebert, Dirk Lehr, Bernhard Sieland, and Matthias Berking. 2016. "Overcome Procrastination: Enhancing Emotion Regulation Skills Reduce Procrastination." *Learning and Individual Differences* 52:10–18.

Ersoy, Fulya. 2023. "Effects of Perceived Productivity on Study Effort: Evidence from a Field Experiment." *Journal of Economic Behavior & Organization* 207:376–391.

Fedyk, Anastassia. 2024. "Asymmetric Naïveté: Beliefs About Self-Control." *Management Science.*

Ferrari, Joseph R., and Dianne M. Tice. 2000. "Procrastination as a Self-Handicap for Men and Women: A Task-Avoidance Strategy in a Laboratory Setting." *Journal of Research in Personality* 34 (1): 73–83.

Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review* 96 (5): 1449–1476.

Giuntella, Osea, Silvia Saccardo, and Sally Sadoff. 2026. "Sleep: Educational Impact and Habit Formation." *Journal of Political Economy* 0 (0): 000–000.

Gneezy, Uri, John A. List, Jeffrey A. Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu. 2019. "Measuring Success in Education: The Role of Effort on the Test Itself." *American Economic Review: Insights* 1 (3): 291–308.

Hall, Crystal C., Jiaying Zhao, and Eldar Shafir. 2014. "Self-Affirmation Among the Poor: Cognitive and Behavioral Implications." *Psychological Science* 25 (2): 619–625.

Hyndman, Kyle, and Alberto Bisin. 2022. "Procrastination, Self-Imposed Deadlines and Other Commitment Devices." *Economic Theory* 74 (3): 871–897.

Hyndman, Kyle, and Alberto Bisin. 2026. "Replication of 'Procrastination, Deadlines, and Performance: Self-Control by Precommitment'." *Psychological Science.*

John, Anett. 2020. "When Commitment Fails: Evidence from a Field Experiment." *Management Science* 66 (2): 503–529.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy* 123 (6): 1227–1277.

Kopylov, Igor. 2012. "Perfectionism and Choice." *Econometrica* 80 (5): 1819–1843.

Kőszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707.

Kőszegi, Botond, George Loewenstein, and Takeshi Murooka. 2022. "Fragile Self-Esteem." *The Review of Economic Studies* 89 (4): 2026–2060.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* 112 (2): 443–477.

Lay, Clarry H. 1986. "At Last, My Research Article on Procrastination." *Journal of Research in Personality* 20 (4): 474–495.

List, John A. 2024. "Optimally Generate Policy-Based Evidence before Scaling." *Nature* 626 (7999): 491–499.

List, John A., Azeem M. Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4): 773–793.

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

———. 2001. "Choice and Procrastination." *The Quarterly Journal of Economics* 116 (1): 121–160.

———. 2008. "Procrastination on Long-Term Projects." *Journal of Economic Behavior & Organization* 66 (2): 161–175.

———. 2015. "Present Bias: Lessons Learned and to Be Learned." *American Economic Review* 105 (5): 273–279.

Peeters, Miranda A. G., and Christel G. Rutte. 2005. "Time Management Behavior as a Moderator for the Job Demand-Control Interaction." *Journal of Occupational Health Psychology* 10 (1): 64–75.

Piedmont, Ralph L., Robert R. McCrae, Rainer Riemann, and Alois Angleitner. 2000. "On the Invalidity of Validity Scales: Evidence from Self-Reports and Observer Ratings in Volunteer Samples." *Journal of Personality and Social Psychology* 78 (3): 582–593.

Romano, Joseph P., and Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4): 1237–1282.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image.* Princeton University Press.

Schraw, Gregory, Theresa Wadkins, and Lori Olafson. 2007. "Doing the Things We Do: A Grounded Theory of Academic Procrastination." *Journal of Educational Psychology* 99 (1): 12.

Schuenemann, Laura, Viviane Scherenberg, Maria von Salisch, and Marcus Eckert. 2022. "'I'll Worry About It Tomorrow' – Fostering Emotion Regulation Skills to Overcome Procrastination." *Frontiers in Psychology* 13.

Shah, Rohan, and Anita Mukherjee. 2026. "Procrastination in Personal Finance: Implications for Estate Planning and Retirement Satisfaction." *Journal of Behavioral and Experimental Finance* 49:101137.

Sirois, Fuschia M., Sisi Yang, and Wendelien van Eerde. 2019. "Development and Validation of the General Procrastination Scale (GPS-9): A Short and Reliable Measure of Trait Procrastination." *Personality and Individual Differences* 146:26–33.

Smithers, Samuel. 2015. "Goals, Motivation and Gender." *Economics Letters* 131:75–77.

Solomon, Laura J., and Esther D. Rothblum. 1984. "Academic Procrastination: Frequency and Cognitive-Behavioral Correlates." *Journal of Counseling Psychology* 31:503–509.

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

Strulov-Shlain, Avner, and Alexandra Steiny Wellsjo. 2026. "Goals, Expectations, and Performance." *Management Science.*

Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112 (S1): S164–S187.

Thiemann, Petra. 2022. "The Persistent Effects of Short-Term Peer Groups on Performance: Evidence from a Natural Experiment in Higher Education." *Management Science* 68 (2): 1131–1148.

Tice, Dianne M, and Roy F. Baumeister. 1997. "Longitudinal Study of Procrastination, Performance, Stress, and Health: The Costs and Benefits of Dawdling." *Psychological Science* 8 (6): 454–458.

Van Lent, Max, and Michiel Souverijn. 2020. "Goal Setting and Raising the Bar: A Field Experiment." *Journal of Behavioral and Experimental Economics* 87:101570.

Xu, Shuai. 2021. "Academic Procrastination of Adolescents: A Brief Review of the Literature." *Psychology and Behavioral Sciences* 10 (6): 198.

A.1 Effect heterogeneity

As preregistered, I estimate heterogeneous effects by gender, interacting the treatment indicators with a female dummy:

$$Y_{it} = \alpha + \sum_k \beta_k \mathbf{1}\{\mathrm{T}_i = k\} + \phi \mathbf{1}\{\mathrm{Female}_i\} + \sum_k \delta_k \Big[\mathbf{1}\{\mathrm{T}_i = k\} \times \mathbf{1}\{\mathrm{Female}_i\}\Big] + \gamma \mathbf{X}_i + \omega_t + \varepsilon_{it},$$

where T_i denotes the treatment arm of student i and $k \in \{\mathrm{Booster}, \mathrm{Goals}, \mathrm{Information}\}$. $\mathbf{X}_i$ is a vector of student-level controls measured at baseline, including separate indicators for Swiss nationality and a Swiss high-school diploma, language of study, and baseline usage of the app. The model includes week fixed effects (ω_t), and standard errors are clustered at the student level.

Figure A.1 reports the implied effects of each treatment for male and female students. The effect of the *Booster* is positive for both groups. It is similar for male and female students for correct and total answers (between 0.03 and 0.04 standard deviations), and larger for female students for distinct questions and time spent (around 0.05 standard deviations, compared to around 0.02 for male students). The *Goals* treatment has no effect on either group. The *Information* treatment has no effect on male students, while its point estimates for female students are positive and larger (around 0.06 standard deviations for correct answers, total answers, and distinct questions answered). Previous interventions in education also find differential effects by gender (Smithers 2015; Clark et al. 2020; Campos-Mercade, Thiemann, and Wengström 2025; Delavande et al. 2026).

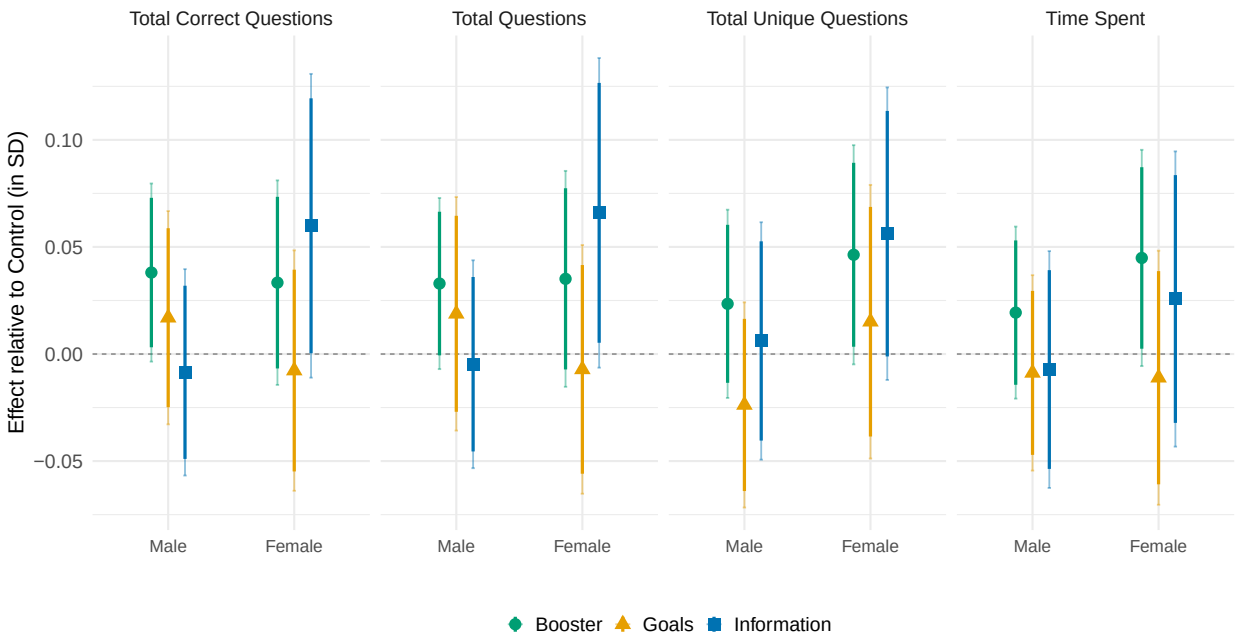


Figure A.1: Effect of Booster, Goals, and Information on weekly app use, separately for male and female students. Each coefficient is the implied OLS estimate relative to the Control group from a regression interacting the treatment indicators with a female dummy, with week fixed effects, controls, and standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals. Outcome variables are standardized relative to the Control group.

By contrast, Figure A.2 shows limited evidence that the effects differ by whether the student obtained a high-school diploma in Switzerland. The point estimates for the *Booster* are similar across the two groups, and the confidence intervals largely overlap for all treatment arms. I preregistered this analysis because of the institutional setting: students from a Swiss high school are admitted to the university automatically, while other students need to pass an entry test. Students from a non-Swiss high school might therefore be more motivated or have higher ability. Swiss nationality and possession of a Swiss high-school diploma enter the regressions as separate indicators.

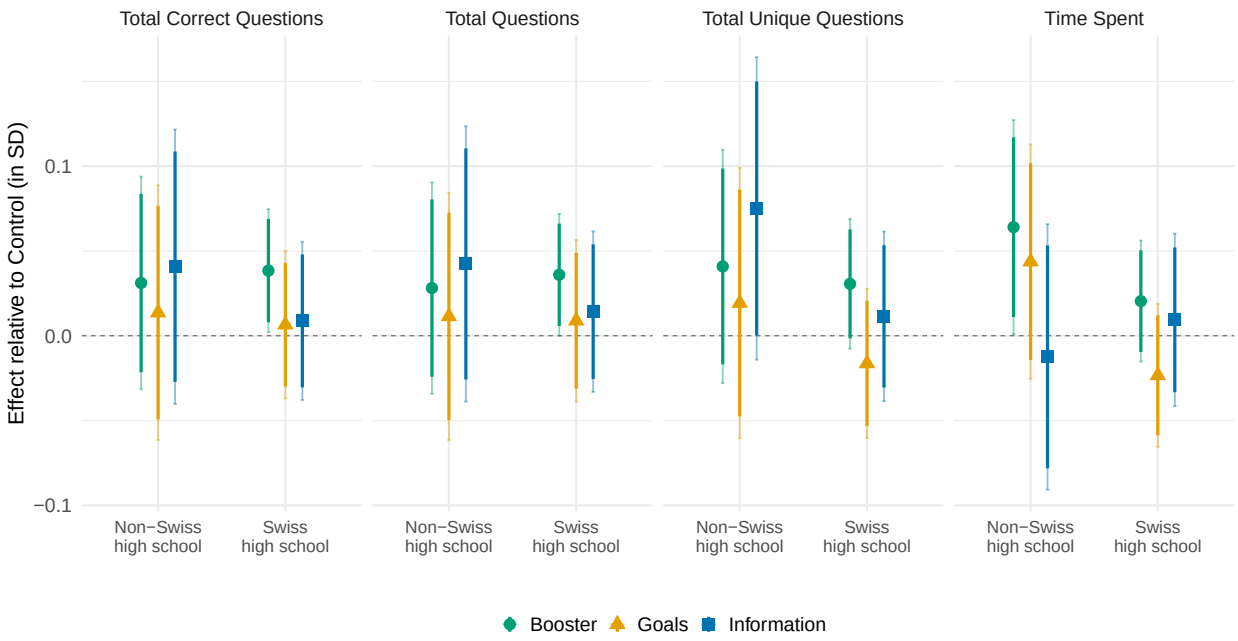


Figure A.2: Effect of Booster, Goals, and Information on weekly app use, separately for students with and without a Swiss high-school diploma. Each coefficient is the implied OLS estimate relative to the Control group from a regression interacting the treatment indicators with a Swiss high-school diploma indicator, with week fixed effects, controls, and standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals. Outcome variables are standardized relative to the Control group.

A.2 Theoretical framework

The theoretical framework connects the traditional β-δ models of time-inconsistent preferences (e.g., Laibson 1997; O'Donoghue and Rabin 1999, 2001, 2008) with recent findings on "ego utility" and fragile self-esteem (e.g., Bénabou and Tirole 2002; Kőszegi 2006; Kopylov 2012; Kőszegi, Loewenstein, and Murooka 2022). For a review of the present-bias literature, see O'Donoghue and Rabin (2015).

A.2.1 Traditional β-δ model

Consider a student facing an intertemporal choice regarding study effort. Assume the student must choose between studying for seven hours on March 1 (Option A) and studying for eight hours on March 14 (Option B). In a standard model of time-inconsistent preferences, the student prefers the "efficient" outcome (Option A) when asked on February 1; however, once March 1 arrives, these preferences reverse, leading them to delay the task. This phenomenon, known as present bias, implies that individuals give greater weight to the present moment relative to any future moment. Formally, this is modeled using (β, δ) preferences. The utility of the student at time t, evaluating a stream of payoffs $(u_t, u_{t+1}, \ldots, u_T)$, is given by

$$U_t = u_t + \beta \sum_{\tau=1}^{T-t} \delta^\tau u_{t+\tau},$$

where $0 < \beta < 1$ captures the present bias (short-run impatience) and $0 < \delta \leq 1$ represents the standard exponential discount factor (long-run patience). For simplicity, assume $\delta = 1$, so that the choice is driven only by present bias rather than by long-run discounting. The parameter β generates the preference reversal: from the perspective of February, both March

1 and March 14 are in the future and are discounted equally by β, so the student chooses the lower cost (seven hours). However, when March 1 arrives, the cost of studying is immediate (u_t) and is not discounted by β, while the future cost on March 14 is discounted by β.

Standard theory distinguishes between naïfs (who are unaware of their future self-control problems) and sophisticates (who are fully aware). A sophisticated student, anticipating their future reversal, would ideally use a commitment device to force themselves to study on March 1.

A.2.2 A β-δ model with self-esteem dependence

Consider now a variation of the decision problem. Before the student decides whether to study on March 1, they receive a phone notification that they failed a previous, unrelated exam. Under the standard β-δ model, preferences are exogenous and stable; this "sunk-cost" news about a past event should not alter their discount factor or their optimal decision about the next exam.

However, there is evidence that decisions depend on the agent's state, in particular on mood and self-esteem (Kőszegi, Loewenstein, and Murooka 2022). Kőszegi (2006) further argues that individuals derive utility not only from consumption or leisure but also from their ego utility: their beliefs about their own ability.

Ego utility gives rise to self-handicapping. Assume the student believes the outcome of an exam depends on both ability (a) and effort (e). A failure after high effort can be interpreted as a signal of low ability ($\downarrow a$), which is psychologically painful. Conversely, a failure after procrastination (low effort) creates a "noisy" signal; the failure can be attributed to lack of

effort rather than lack of ability, which protects the student's self-image.

I model this mechanism in reduced form by letting the present-bias parameter of the traditional β-δ model depend on the agent's current state of self-esteem, denoted by s. The utility function becomes

$$U_t(s) = u_t + \beta(s) \sum_{\tau=1}^{T-t} \delta^\tau u_{t+\tau}.$$

Here, $\beta(s)$ is the "self-esteem-dependent" present bias.[2] I assume that $\beta'(s) > 0$. As s increases, $\beta(s) \to 1$: a student with high self-esteem feels secure in their ability, has less need to protect their ego by self-handicapping, and behaves closer to a standard exponential discounter. As s decreases, $\beta(s)$ falls. The lower β captures in reduced form the pull toward delay created by the immediate value of "protecting the ego" (making the signal of ability noisier), and this increase in impatience can lead to procrastination.

Following the notion of fragile self-esteem in Kőszegi, Loewenstein, and Murooka (2022), I assume that s takes one of two values: a confident state $\bar{s}$ and a fragile state $\underline{s}$. A negative shock, such as the failed exam notification, moves s from $\bar{s}$ to $\underline{s}$, and present bias rises from $\beta(\bar{s})$ to $\beta(\underline{s})$. The cost of immediate effort now outweighs the discounted future benefits because of the immediate "ego cost" of potentially proving oneself a low-ability type. Contrary to the standard model, an exogenous shock to self-esteem therefore exacerbates present bias, and self-handicapping procrastination becomes the agent's optimal strategy to preserve their remaining self-esteem.

The model predicts that boosting self-esteem should shift some agents from $\beta(\underline{s})$ to $\beta(\bar{s})$

2. Augenblick, Niederle, and Sprenger (2015) also show that β is not a fixed trait: the same individuals exhibit present bias in the domain of real effort, but not in the domain of monetary payoffs.

and reduce their propensity to procrastinate. By contrast, providing the same information on past study time without a positive message should not shift s, and should therefore not reduce procrastination through this channel.

A.3 Tables

Table A.1: Balance Table: students in the experimental sample vs. other students

Variable	Experimental Sample	Other Students	Diff.
Female	0.37	0.36	0.01
	(0.01)	(0.02)	(0.02)
Age	19.77	19.74	-0.05
	(0.03)	(0.06)	(0.07)
Swiss	0.75	0.69	0.07***
	(0.01)	(0.02)	(0.02)
German Track	0.65	0.51	0.16***
	(0.01)	(0.02)	(0.02)
Brian User	0.95	0.27	0.71***
	(0.00)	(0.02)	(0.02)
Total Questions	762.82	86.60	647.59***
	(17.41)	(12.17)	(22.50)
Total Correct Questions	547.58	63.00	465.78***
	(13.34)	(9.72)	(17.52)
Total Unique Questions	409.44	47.04	350.95***
	(8.07)	(5.16)	(10.34)
Time Spent (minutes)	665.68	106.99	530.35***
	(23.16)	(18.10)	(29.13)
N	2774	826	

This table reports means and standard errors (in parentheses) for baseline variables for individuals in the experiment and individuals not in the experiment. The variables refer to the period before the randomization. Only students that had downloaded the app and had notifications enabled were included in the experiment. The difference and its standard error (heteroskedasticity-robust) are estimated controlling for experiment wave fixed effects. * p<0.1, ** p<0.05, *** p<0.01.

Table A.2: Summary statistics for experimental sample

Variable	N	Mean	SD
Female	2774	0.37	0.48
Age	2774	19.77	1.44
Swiss	2774	0.75	0.43
German Track	2774	0.65	0.48
Brian User	2774	0.95	0.22
Total Questions	2774	762.82	917.21
Total Correct Questions	2774	547.58	702.65
Total Unique Questions	2774	409.44	425.02
Time Spent (minutes)	2774	665.68	1219.88
Procrastination	482	25.76	6.31
Self-Esteem	482	38.51	6.58
Time Management	477	34.62	6.06
Expected Grade (highest=6)	476	4.91	0.50

This table reports summary statistics for baseline variables for individuals in the experiment. The variables refers to the period before the randomization. Procrastination, self-esteem, time management, and the expected grade in Business Administration A refers to the baseline survey. Time spent is the total time in the Business Administration A and Economics apps before the first notification, in minutes, winsorized at the 99th percentile. Students that failed the attention check in the baseline survey or that did not answer the questions on procrastination and self-esteem were excluded from the sample.

Table A.3: Balance Table: students in the experimental sample by survey wave

Variable	Baseline	Midline	Endline	P value (F test)
Female	0.45	0.48	0.48	0.423
	(0.02)	(0.03)	(0.03)	
Age	19.56	19.70	19.63	0.318
	(0.06)	(0.09)	(0.08)	
Swiss	0.72	0.78	0.74	0.165
	(0.02)	(0.03)	(0.03)	
German Track	0.70	0.60	0.57	0.000***
	(0.02)	(0.03)	(0.03)	
Brian User	0.94	0.95	0.97	0.135
	(0.01)	(0.01)	(0.01)	
Total Questions	902.76	922.31	901.30	0.770
	(47.80)	(60.79)	(57.09)	
Total Correct Questions	655.28	656.70	644.61	0.822
	(36.64)	(46.47)	(43.97)	
Total Unique Questions	466.27	504.70	501.64	0.426
	(21.43)	(28.70)	(26.32)	
Time Spent (minutes)	660.78	746.54	679.99	0.571
	(53.72)	(94.21)	(74.78)	
Treatment: Control	0.30	0.35	0.31	0.369
	(0.02)	(0.03)	(0.03)	
Treatment: Booster	0.38	0.33	0.36	0.334
	(0.02)	(0.03)	(0.03)	
Treatment: Information	0.14	0.15	0.05	0.033**
	(0.02)	(0.02)	(0.01)	
Treatment: Goals	0.18	0.17	0.28	0.281
	(0.02)	(0.03)	(0.03)	
N	482	227	246	

This table reports means and standard errors (in parentheses) for baseline variables for individuals in the experiment by survey wave. The variables refer to the period before the randomization. The last column reports the p-value of an F-test that the means are equal across survey waves, controlling for experiment wave fixed effects, with standard errors clustered at the student level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.4: Effects of Treatment on Weekly Outcomes

	Total Correct Questions	Total Questions	Total Unique Questions	Time Spent
Booster	0.036**	0.034**	0.032*	0.029*
	(0.016)	(0.016)	(0.017)	(0.016)
	[0.048]	[0.070]	[0.100]	[0.100]
Goals	0.008	0.009	−0.009	−0.010
	(0.020)	(0.021)	(0.020)	(0.019)
	[0.898]	[0.898]	[0.898]	[0.898]
Information	0.015	0.020	0.024	0.005
	(0.021)	(0.022)	(0.023)	(0.023)
	[0.682]	[0.556]	[0.504]	[0.850]
N (students)	2,774	2,774	2,774	2,774
N (weeks)	58	58	58	58
Week FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Clustered SEs	Yes (Student)	Yes (Student)	Yes (Student)	Yes (Student)

* p <0.1, ** p <0.05, *** p <0.01

Each column reports OLS estimates of the effect of each treatment arm relative to Control on weekly app use, pooling both waves. The sample includes all weeks of the Business Administration A and Economics apps from the first notification (10 November 2024 and 9 November 2025) to the last Sunday before the exam (19 January 2025 and 18 January 2026), and all weeks of the Business Administration B app in the following semester up to the last Sunday before the exam (22 June 2025 and 21 June 2026). All regressions include week fixed effects and control for student gender, nationality, language of study, and baseline usage of the app. Time spent is the weekly time in the app in minutes, winsorized at the 99th percentile separately for each course and wave. Outcome variables are standardized relative to the Control group (weekly mean and standard deviation of the Control group), separately for each course and wave. Standard errors clustered at the student level in parentheses. Romano-Wolf step-down p-values (Romano and Wolf, 2005; see also List, Shaikh and Xu, 2019), adjusted for testing the four outcomes and computed separately for each treatment arm from 1,000 bootstrap draws clustered at the student level, in square brackets. Stars refer to the unadjusted p-values.

A.4 Figures



Figure A.3: Brian's interface showing the course page, an example of a question, and the home page.

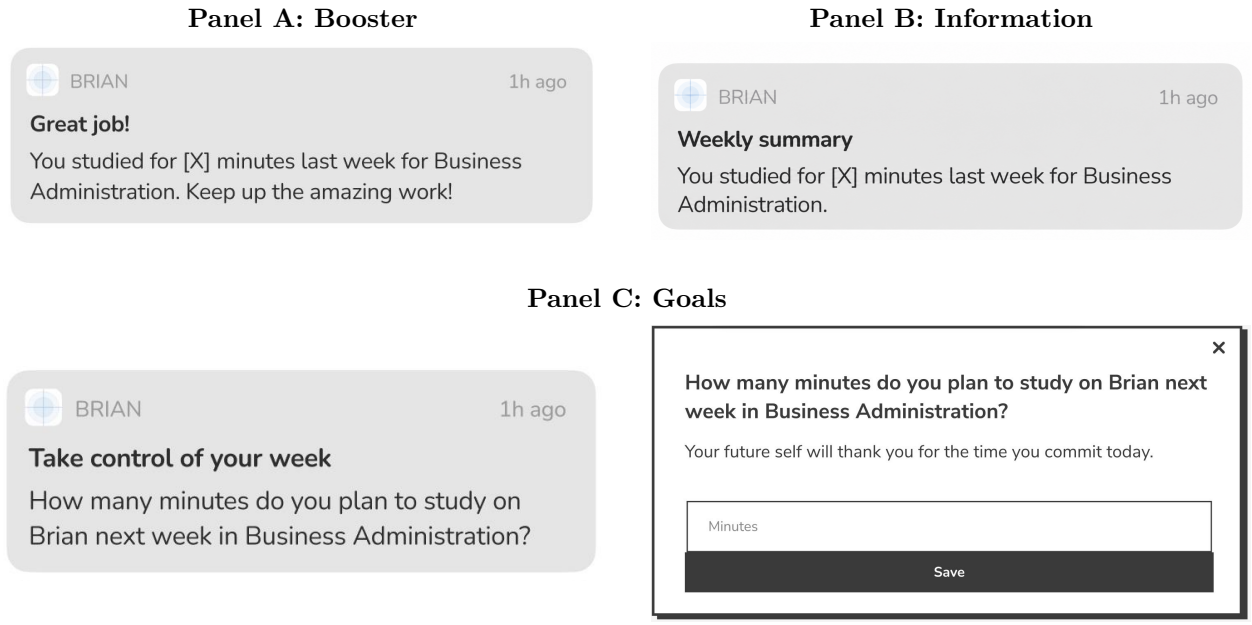


Figure A.4: Text of the notifications of the Booster (Panel A), Information (Panel B), and Goals (Panel C) treatments. In the Goals treatment, clicking on the notification opens an input box in the app (Panel C, right).

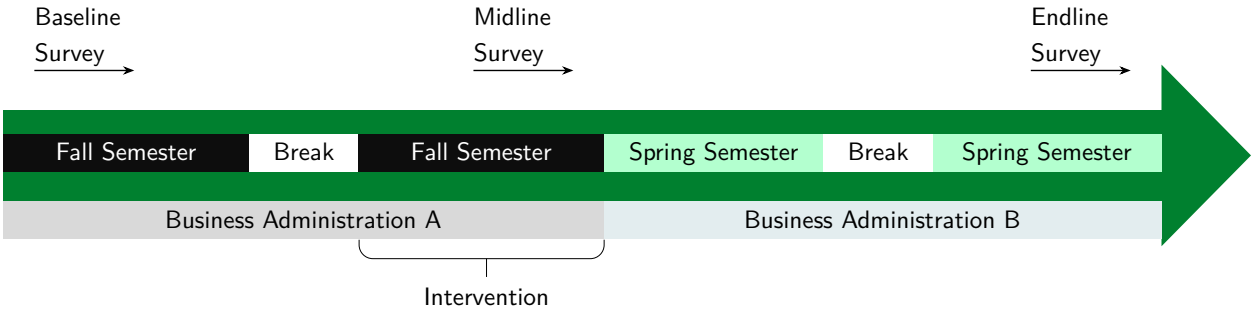


Figure A.5: Timeline of the experiment.

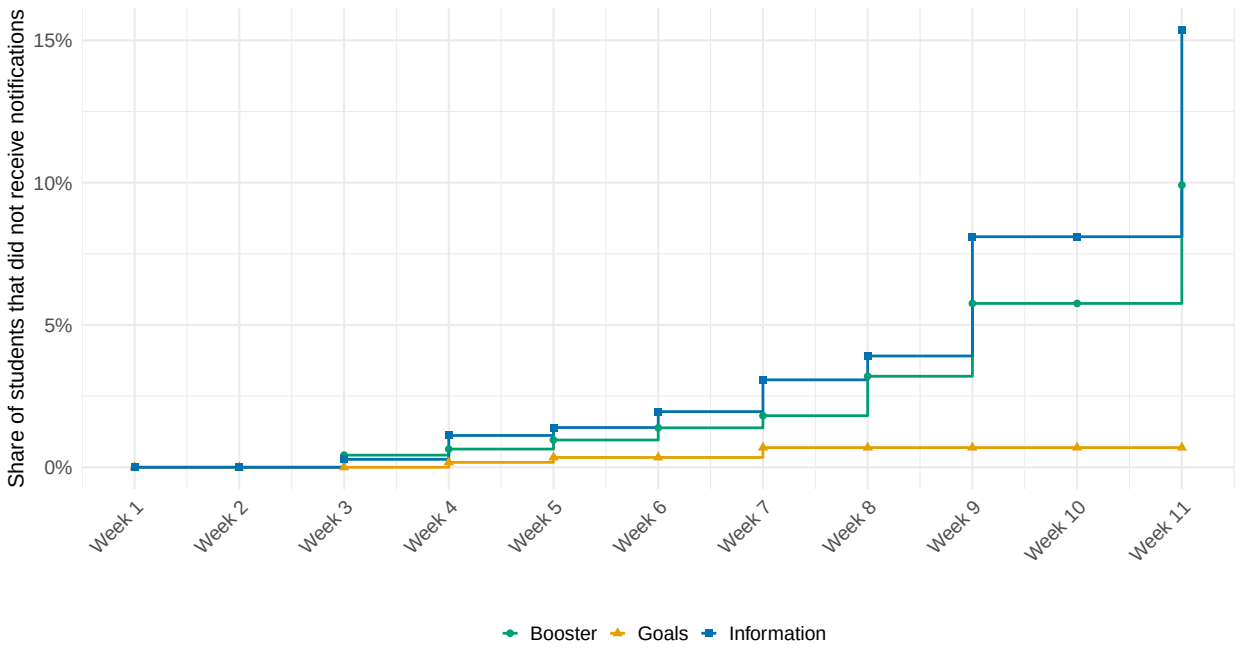


Figure A.6: Share of students in each treatment arm who did not receive the notification because they had disabled notifications or uninstalled the app, by week of the intervention.



Figure A.7: Share of students in the Booster group with zero app use (who did not receive that week's notification) and with low app use (positive, but under five minutes), by week of the intervention.

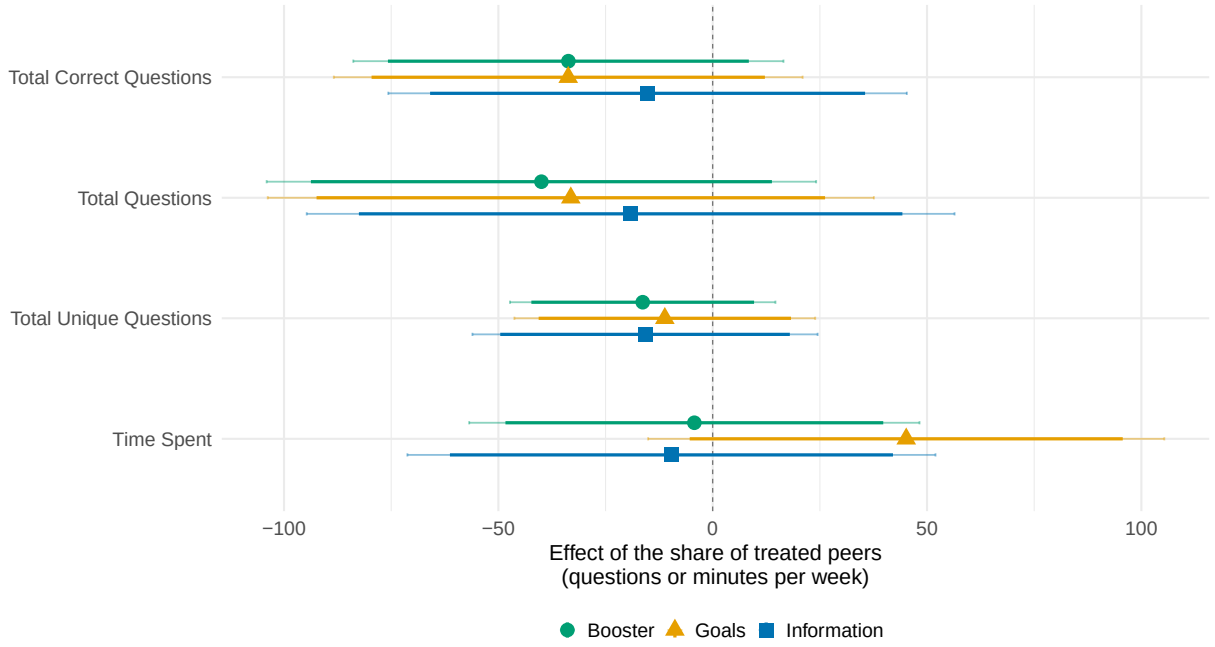


Figure A.8: Effect of the share of peers in each treatment arm within the orientation week group on weekly app use (questions or minutes per week). Each coefficient is the OLS estimate from a regression controlling for the student's own treatment assignment, with week fixed effects, controls, and standard errors clustered at the student level. Thick and thin bars represent 90% and 95% confidence intervals.

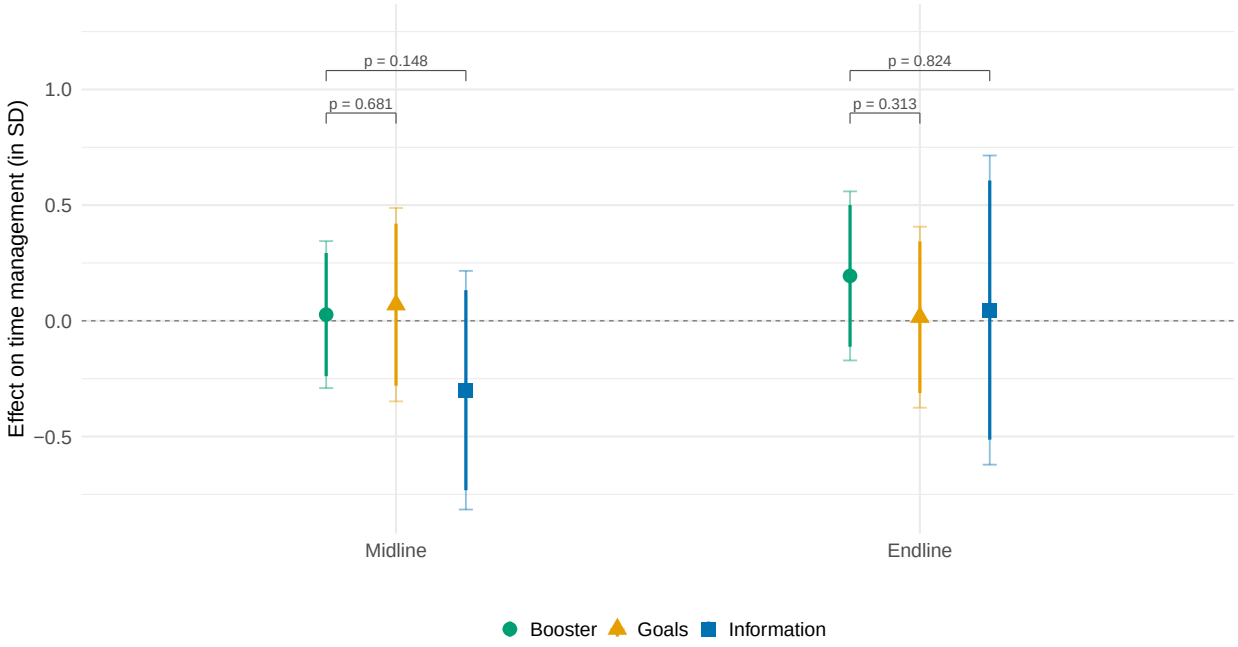


Figure A.9: Effect of Booster, Goals, and Information on reported time management in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Brackets report the p-values of the tests of equality between the Booster coefficient and the Goals and Information coefficients. Time management is measured using the Short Time Management Behaviors Scale (Peeters and Rutte 2005) and standardized.

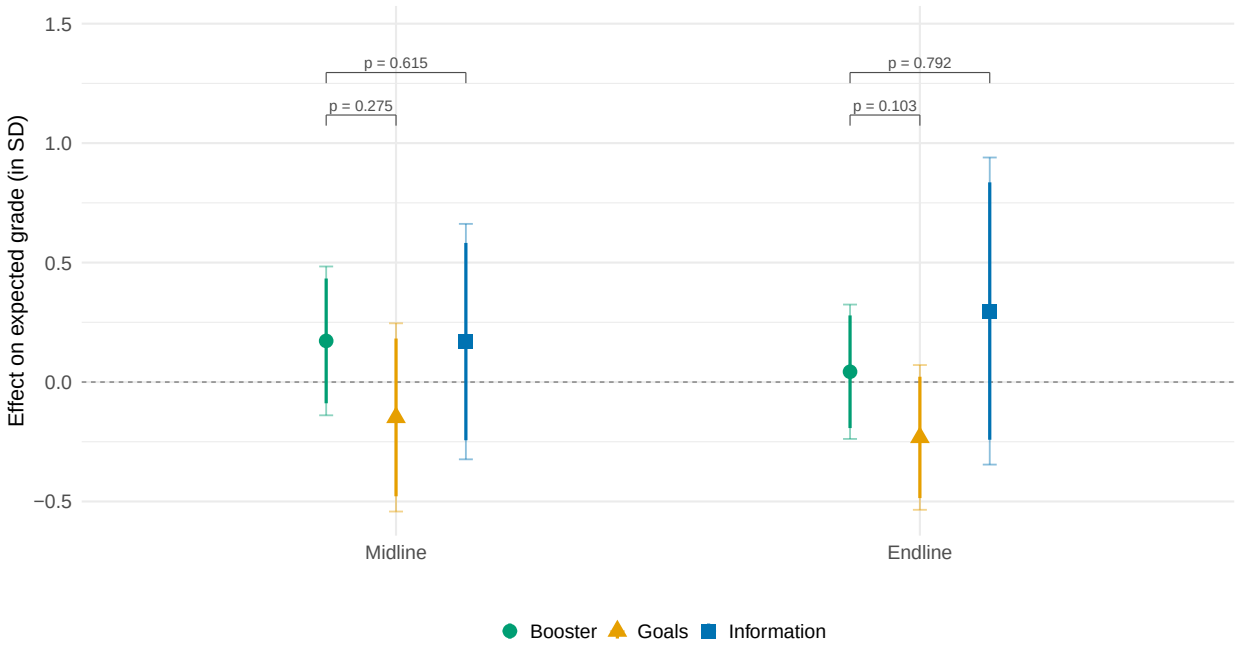


Figure A.10: Effect of Booster, Goals, and Information on the expected grade in Business Administration in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Brackets report the p-values of the tests of equality between the Booster coefficient and the Goals and Information coefficients. The expected grade is standardized.

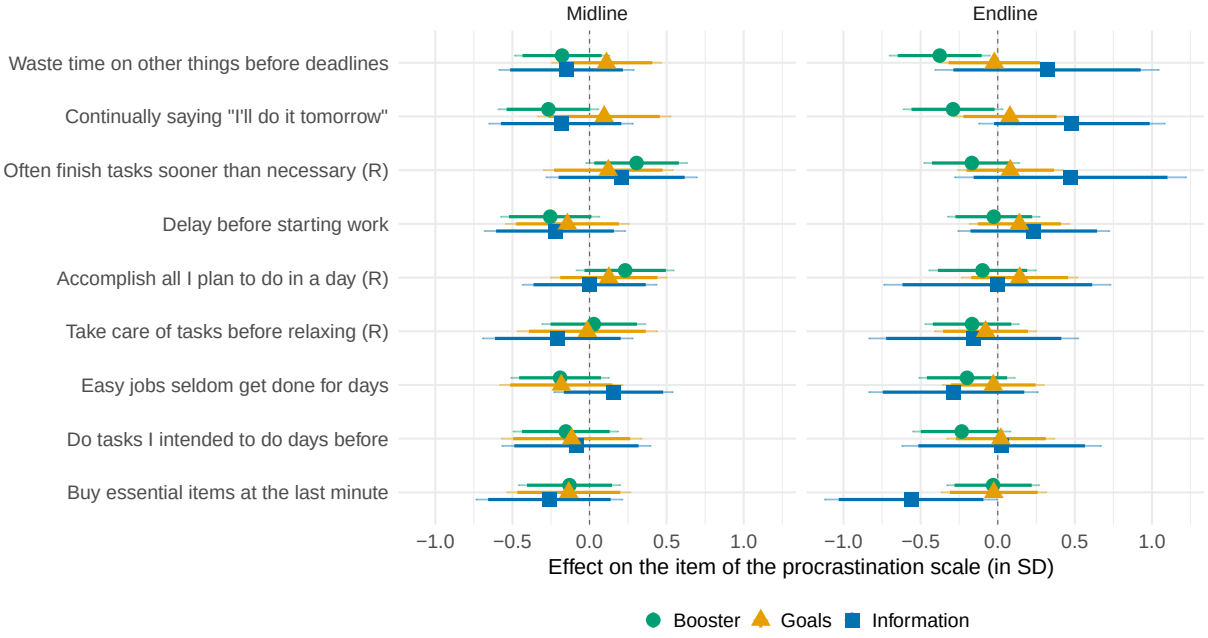


Figure A.11: Effect of Booster, Goals, and Information on each item of the Short General Procrastination Scale (Sirois, Yang, and van Eerde 2019) in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Items marked (R) are reverse-coded, so that higher values always indicate more procrastination. Each item is standardized. The full text of the items is in Section A.5.

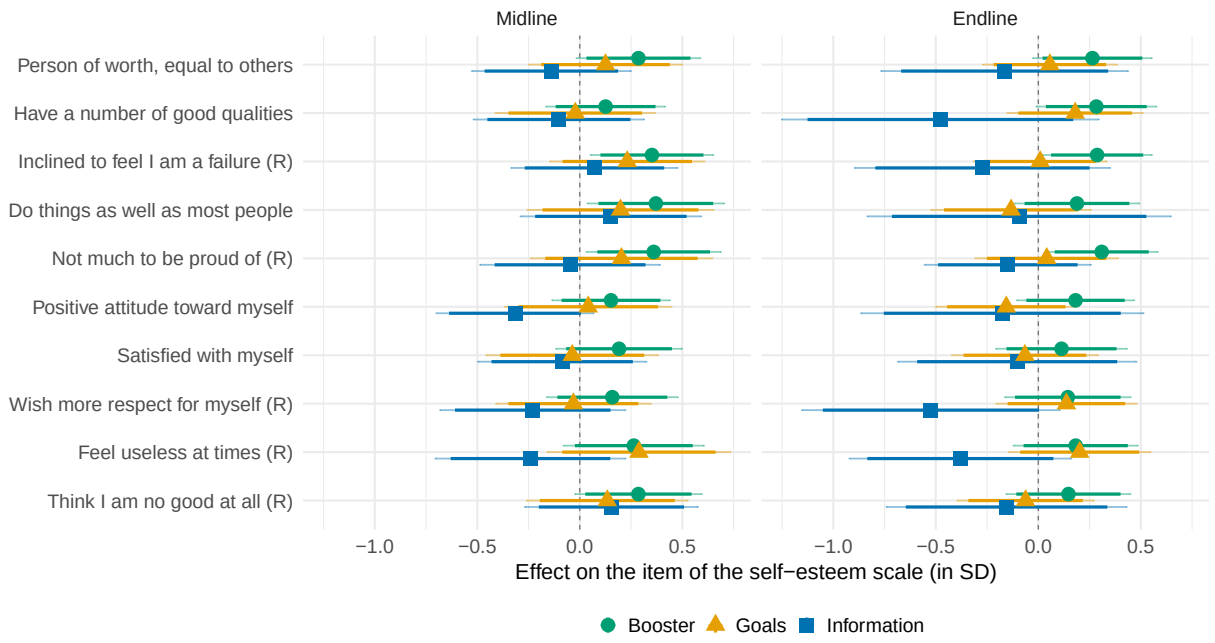


Figure A.12: Effect of Booster, Goals, and Information on each item of the Rosenberg Self-Esteem Scale (Rosenberg 1965) in the midline and endline surveys. Each coefficient is the OLS estimate relative to the Control group, with robust standard errors. Thick and thin bars represent 90% and 95% confidence intervals. Items marked (R) are reverse-coded, so that higher values always indicate higher self-esteem. Each item is standardized. The full text of the items is in Section A.5.

A.5 Surveys

This appendix reports the text of the survey. The order of the questions related to procrastination, self-esteem, time management, and the attention check was randomized. Self-esteem is measured using the Rosenberg Self-Esteem Scale (Rosenberg 1965); procrastination is measured using the Short General Procrastination Scale (Sirois, Yang, and van Eerde 2019); time management is measured using the Short Time Management Behaviors Scale (Peeters and Rutte 2005).

- **Procrastination**: On a scale from 1 to 5, where 1 means "Strongly Disagree" and 5 means "Strongly Agree", please indicate your level of agreement with the following statements:
  - In preparing for some deadlines, I often waste time by doing other things
  - I am continually saying "I'll do it tomorrow"
  - I often have a task finished sooner than necessary
  - I generally delay before starting work I have to do
  - I usually accomplish all the things I plan to do in a day
  - I usually take care of all the tasks I have to do before I settle down and relax for the evening
  - Even with jobs that require little else except sitting down and doing them, I find they seldom get done for days
  - I often find myself performing tasks that I had intended to do days before

  - I usually buy even essential items at the last minute

- **Self-Esteem**: On a scale from 1 to 5, where 1 means "Strongly Disagree" and 5 means "Strongly Agree", please indicate your level of agreement with the following statements:
  - I feel that I am a person of worth, at least on an equal plane with others
  - I feel that I have a number of good qualities
  - All in all, I am inclined to feel that I am a failure
  - I am able to do things as well as most other people
  - I feel I do not have much to be proud of
  - I take a positive attitude towards myself
  - On the whole, I am satisfied with myself
  - I wish I could have more respect for myself
  - I certainly feel useless at times
  - At times I think I am no good at all

- **Time Management**: On a scale from 1 to 5, where 1 means "Seldom true" and 5 means "Very often true", please indicate the extent to which the following statements apply to you:
  - I feel in control of my time
  - I review my daily activities to see where I am wasting my time
  - I break complex, difficult projects down into smaller manageable tasks
  - I set short-term goals for what I want to accomplish in a few days or weeks

  - I set deadlines for myself when I set out to accomplish a task
  - I look for ways to increase the efficiency with which I perform my work activities
  - I set priorities to determine the order in which I will perform tasks each day
  - I review my goals to determine if they need to be adjusted
  - I finish top-priority tasks before going on to less important ones
  - I make a list of things to do each day and check off each task as it is accomplished

- This is a question to check whether you are paying attention and reading the questions carefully. Please select both "strongly disagree" and "strongly agree" to continue
- What grade do you expect to receive in Business Administration? Please report a number from 1 to 6
- On a scale from 1 to 5, where 1 means "Not at all certain" and 5 means "Extremely certain", how certain are you about your previous answer?
- In relation to your grade in the Business Administration course, how do you think you will rank compared to others in your cohort?
- On average, how many minutes per day do you spend studying Business Administration? Please provide a specific number.