Performance Incentives in Education: The Role of Goal Mismatch

Campos-Mercade, Pol; Thiemann, Petra; Wengström, Erik

2025

Document Version:
Other version

Working Paper 2025:5

Department of Economics
School of Economics and Management

Performance Incentives in Education: The Role of Goal Mismatch

Pol Campos-Mercade
Petra Thiemann
Erik Wengström

May 2025

LUND UNIVERSITY

Performance Incentives in Education: The Role of Goal Mismatch*

Pol Campos-Mercade, Petra Thiemann, Erik Wengström

May 7, 2025

Abstract

We conduct a field experiment studying how financial incentives for achieving specific course grades affect university students, whether effects vary by ability, and whether allowing students to choose their goals improves outcomes. We find that incentives negatively affect performance, particularly among low-ability students assigned high goals. Survey data suggest this negative impact arises from a mismatch between assigned goals and students' abilities, distorting their reference points and expectations. Allowing students to choose their goals partially mitigates this effect but does not eliminate it. Our results caution against incentives in education and highlight a novel mechanism through which incentives can backfire.

Keywords: Incentives, Performance goals, Academic performance, Field experiment.

JEL Codes: C93, D90, I22, I23.

*Pol Campos-Mercade, Lund University and Institute for Future Studies, pol.campos@nek.lu.se; Petra Thiemann, Lund University and IZA, petra.thiemann@nek.lu.se; Erik Wengström, Lund University and Hanken School of Economics, erik.wengstrom@nek.lu.se. This paper was presented at the Berlin Behavioral Economics Seminar, TrygFonden at Aarhus University, the 2022 CEN workshop at Copenhagen Business School, the 2023 IZA Workshop on the Economics of Education, Lucerne University, the University of Bamberg, IAB Nuremberg, and Lund University. We thank the audiences for their valuable input. We also thank Jordi Brandts, Andrés Gago, Uri Gneezy, Víctor González-Jímenez, Simone Häckl, Devin Pope, Pedro Rey-Biel, Dan-Olof Rooth, Florian Schneider, Charlie Sprenger, Robert Stüber, Mattie Toma, Roel van Veldhuizen, Marie-Claire Villeval, Georg Weizsäcker, and Andrej Woerner for helpful comments and discussions. This project received funding from the Crafoord Foundation. The analysis was pre-registered at the AEA RCT registry (AEARCTR-0003507). The experiment was evaluated by the Regional Ethics Review Board in Lund on August 15, 2018 (Regionala Etikprövningnämnden, Dnr 2018/609). The board determined that the project was exempt, i.e., did not require approval based on the Swedish Ethics Review Act. Moreover, in an advisory opinion, the board also stated that there are no ethical obstacles to the project proceeding.

1 Introduction

Incentives to achieve performance goals are pervasive across a wide range of domains, including business, public health, and education (Gneezy et al. 2011). These incentives can take many forms: bonuses for employees who meet sales quotas (Lazear 2000a; Ockenfels et al. 2014; Friebel et al. 2017), rewards for individuals who follow health recommendations (Charness and Gneezy 2009; Lacetera et al. 2013; Campos-Mercade et al. 2021), and scholarships or awards for students who reach grade targets (Angrist et al. 2009; Scott-Clayton 2011; Montalbán 2023).

A large theoretical literature suggests that incentives can influence individuals' performance through a variety of channels. They may increase effort directed toward reaching the performance goal (e.g., Becker 1967; Lazear 2000b), but they may also affect intrinsic motivation (Deci and Ryan 1985; Gneezy and Rustichini 2000; Bénabou and Tirole 2003; Prendergast 2008) or move individuals' reference points (Kőszegi and Rabin 2006; Abeler et al. 2011; Koch and Nafziger 2016; O'Donoghue and Sprenger 2018). Importantly, most models predict that the effects of incentives will crucially depend on individuals' performance in the absence of incentives—often referred to as their *ability*. This implies that a given performance goal may affect individuals differently, potentially motivating some individuals while discouraging others. For example, a high target may be effective for a high-ability individual but counterproductive for someone with lower ability. If such heterogeneity in responses exists, then allowing individuals to select their own goal may be an effective policy tool to improve performance.

The impacts of incentivized goals and of goal choice have been extensively studied theoretically (Koch and Nafziger 2011; Golman and Loewenstein 2012; Gomez-Minambres 2012; Hsiaw 2013; Koch and Nafziger 2016; González-Jiménez 2024) and in laboratory experiments (Corgnet et al. 2015; Corgnet et al. 2018; Goerg et al. 2019; González-Jiménez et al. 2019; Koch and Nafziger 2020; Campos-Mercade et al. 2022). However, little is known about how individuals respond to incentivized goals—and the opportunity to select them—in field settings, where behavior may not align with theory and laboratory settings. Addressing this gap is particularly important in education, where incentives such as scholarships, awards, and grade targets are pervasive and may shape students' long-term academic trajectories and career outcomes.

This paper addresses both the broader gap regarding field evidence and the specific gap in educational contexts by empirically examining the effects of incentivized goals and of goal choice in a real-world educational environment.

We conduct a pre-registered field experiment that incentivized students to reach specific performance-based goals in a seven-week full-time university course. The setting is ideal as all students took another similar course prior to our experiment, providing us with a baseline measure of ability that allows us to understand how different individuals react to different performance goals. In addition, since all students were already enrolled to this full-time course and did not have the opportunity to strategically alter their course choices, our setting ensures that all participants were held to the same performance standards irrespective of treatment. Finally, we collected a very rich set of additional data—before, during, and after the course—that helps us identify potential mechanisms.

Our sample consists of 425 students enrolled in the "Introduction to Economics" course sequence at Lund University, one of Sweden's largest universities, which is taught full-time over one semester. The course sequence starts with a course in Microeconomics, where we measure students' baseline performance. Subsequently, all students move into the Macroeconomics course, where the randomized intervention takes place. Our experimental design randomizes students into one control group and three treatment groups. The "exogenous-low" treatment group offers students SEK 2,000 ($220 at the time) if their Macroeconomics grade is C or higher. The "exogenous-high" treatment group offers students SEK 4,000 ($440) if their Macroeconomics grade is an A. Finally, the "endogenous" treatment group allows students to choose whether they want the *low goal* (earning SEK 2,000 if they reach a grade of C or higher) or the *high goal* (earning SEK 4,000 if they reach a grade of A).[1]

We find that offering such substantial financial incentives to achieve performance goals harms students' performance. Students in the incentive treatments performed statistically significantly worse than those in the control group, by 0.20 standard deviations on average (main pre-

[1]To provide two benchmarks for the size of the incentives, a full student government grant is approximately SEK 4,000 per month, meaning that the payment of the high goal corresponds to about the same as a students' monthly allowance (in practice, many students combine the government grant with loans to cover their costs). Student jobs pay about SEK 100–110 per hour, meaning that the payment of the high goal corresponds to 40 hours of work in a student job.

registered test). This effect is driven primarily by students who are assigned to the exogenous-high treatment group, who perform 0.30 standard deviations worse. While participants in the exogenous-low and endogenous treatment groups also perform worse than those in the control group, these differences are smaller and not statistically significant.

These results hide notable heterogeneities. By classifying students into ability groups using their grades in the previous course, we find that the negative impact of the exogenous-high treatment is concentrated among low-ability students. This finding suggests that a potential misalignment between performance goals and ability may drive the adverse effects of the incentives. An exploratory analysis confirms this hypothesis and shows that "mismatched" students—low-ability students who received the high goal and high-ability students who received the low goal—performed 0.32 standard deviations worse than their control counter-parts, driving the negative treatment effects of the incentives.[2]

Our detailed survey allows us to dig deeper into understanding these results. In the survey, before the treatment assignment, we allowed students in all groups to choose whether they would prefer the low or the high goal. We find that the adverse effects are concentrated among students who received a goal that did not match their preferences. Students who did not receive their preferred goal performed 0.36 standard deviations worse than students in the control group. These results are mainly driven by students who preferred a low goal at baseline and received a high goal instead.

A potential mechanism is that students have their own reference point before receiving any incentive and then shift their target upward to a less attainable one after being assigned a performance goal. At baseline and each week throughout the experiment, we elicited students' target grades (desired course grades). Initially, we find that students set ambitious but attainable target grades, in line with theories on optimal setting of goals as reference points (Heath et al., 1999, Koch and Nafziger, 2016, 2020, Clark et al., 2020, González-Jiménez, 2024). However, after receiving the incentive, students update their target grade: those with the exogenous-high

[2]In addition to ability, we also pre-registered that we would study gender differences, without any clear hypothesis, as previous research often finds that male and female students react differently to treatments (e.g., Angrist et al. 2009, 2014a and Campos-Mercade and Wengström 2020). We find that the negative treatment effects are driven by males, who perform 0.28 standard deviations worse when assigned incentives. The effect among females is much smaller (0.01 standard deviations) and not statistically significant.

goal are more likely to set their goal at an A rather than at a B or a C. This new reference point may be suboptimal for their performance, and we observe the most significant negative effects among students with such misaligned goals.

These results suggest that allowing students to choose their own goals could mitigate the negative effects of incentives by aligning goals with individual preferences and abilities. Consistent with this, we find weaker and statistically insignificant effects for students in the endogenous treatment group, where students chose between the low and high goals. However, the point estimates remain negative, suggesting that self-chosen goals do not fully eliminate the detrimental impacts of incentives. One possible explanation is that individuals may not be well equipped to choose optimal incentivized goals. For example, some low-ability students may overestimate their ability and select an overly ambitious target, while other high-ability students may shy away from more difficult goals due to risk aversion. Together, these findings highlight that while choice helps, it cannot fully overcome the negative effects of mismatched incentives.

Our results are robust to concerns about both selection into the exam and spillover effects between students. Incentives did not affect whether students chose to take the exam, and reweighting our estimates by students' predicted likelihood of exam participation yields similar results. We also find no evidence that spillovers from treated to control students impact our findings: estimates remain stable when controlling for peer connections, controlling for the treatment assignment of peers, and splitting the sample by students' level of social interaction. Moreover, aggregate yearly performance trends and survey responses suggest no evidence of spillovers to the control group. These checks support the interpretation that our estimates capture direct effects of the interventions.

In sum, this paper reports three main findings. First, we document that incentivized performance goals negatively affect student performance. Second, we find that the negative effects are concentrated among students whose performance goals are mismatched with both their ability and preferences. Third, we find that allowing students to choose their own goals does not fully mitigate the negative incentive effects.

Our paper contributes to a large literature on the effects of incentives on performance. Standard economic theory makes clear predictions on how incentives affect performance (e.g., Becker 1967 and Lazear 2000a), and the evidence sometimes finds patterns in line with these predictions (e.g., Lazear 2000a; Charness and Gneezy 2009; DellaVigna and Pope 2017; Englmaier et al. 2024). However, the behavioral effects of incentives have also often turned out to be complex and even counterintuitive (e.g., Bénabou and Tirole 2003; Prendergast 2008), attracting a wide range of empirical work. For example, the review by Deci et al. (1999) shows that extrinsic rewards can sometimes undermine people's intrinsic motivation. Similarly, Gneezy et al. (2011) review existing evidence on incentives in education, pro-social actions, and health, finding that incentives sometimes backfire and highlighting the importance of understanding *when* they backfire and *why* (see also Bowles and Polania-Reyes 2012).

Our paper provides the first field evidence that incentivized performance goals can backfire when they are mismatched with students' abilities and aspirations. We argue that incentivized performance targets may act as reference points, and mismatched targets can negatively impact performance. This empirical result is consistent with a largely theoretical literature highlighting how incentivized goals shape individuals' reference points and, consequently, affect their behavior (De Meza and Webb 2007; Dittmann et al. 2010; Herweg et al. 2010; Gomez-Minambres 2012; Corgnet et al. 2015; González-Jiménez et al. 2019; González-Jiménez 2024). Our field evidence thus contributes directly to the understanding of why and under which conditions incentives may lead to unintended consequences.

Second, we contribute to a literature that studies how allowing individuals to set their own goals affects performance (Heath et al. 1999; Koch and Nafziger 2011; Goerg and Kube 2012; Hsiaw 2013; Koch and Nafziger 2016; Herranz-Zarzoso and Sabater-Grande 2018; González-Jiménez et al. 2019; Clark et al. 2020; Koch and Nafziger 2020). This literature typically finds that self-chosen goals act as reference points, and allowing people to choose them typically increases performance. For example, Koch and Nafziger (2016) and Koch and Nafziger (2020) show both theoretically and empirically how individuals can set goals to fight self-control problems and increase their performance. Similarly, Clark et al. (2020) show that allowing students

to choose task-based goals improves performance.[3] Our paper complements this literature by studying a field setting in which individuals choose payment-relevant goals. Our findings that mismatched goals can reduce performance indicate that it may often be preferred to allow individuals to choose their own goals in accordance with their ambitions. However, we interpret our results cautiously: in our experiment, individuals may not always choose optimal goals for their performance, and even self-chosen goals may have negative effects.

Finally, our paper adds to the extensive literature on the role of financial incentives on student performance at different levels of schooling (Angrist et al. 2009; Angrist and Lavy 2009, Fryer 2011; Bettinger 2012; Behrman et al. 2015; De Paola et al. 2012; Levitt et al. 2016; Bellés-Obrero 2020; Montalbán 2023).[4] In particular, we add to a set of papers that study how incentives affect students' performance. Leuven et al. (2010) document a positive effect of a threshold incentive for passing the first year at university on high-ability students, but a negative effect on low-ability students. Similarly, Campos-Mercade and Wengström (2020) find positive effects of a threshold financial incentive only for marginal students below the targeted grade, and negative longer-run effects on low-ability students who were further away from the target. Levitt et al. (2017) find positive effects for high-school students around the performance threshold. These papers study the effects of a single performance goal typically aimed at the median student. Beyond comparing the effects of exogenous and endogenous goals, we complement this work by testing how goals of different difficulty—in particular, clearly low and high goals—impact students on different parts of the performance distribution as well as students with different aspirations.

[3] See also Woerner et al. (2024) and Campos-Mercade et al. (2024b) for studies in which individuals choose the type of incentives they receive for engaging in mindfulness practice and vaccination.

[4] A major challenge of much of this literature is that students who are offered incentives may opt for easier courses or programs, hence potentially biasing upwards the effects of incentives on performance (Casey et al. 2018, Cornwell et al. 2005). We add to this work by studying a setting where students' performance in all treatments can be directly compared.

2 Background and design

2.1 Institutional background

The experiment was conducted during the 2018–2019 academic year with undergraduate students at Lund University's School of Economics and Management. Admission to the program is competitive, handled through Sweden's centralized admission system, and attracts both Swedish and international students.[5] All students at the School of Economics and Management must complete the course sequence "Introduction to Economics" within their first study year. The course sequence is also open for students in other social science programs. It is offered twice a year, once in the Spring and once in the Fall semester.

The course sequence is to be studied full-time for 18 weeks in total. It consists of four courses taught in order: first, Microeconomics over seven weeks (full-time); then, Macroeconomics over seven weeks (full-time); and then, Finance and International Economics over four weeks (both courses adding up to full-time studies). For each course, students get a letter grade ranging from F (fail) to A (best) based on the number of points they obtained in the course (out of 100).[6] The examination takes place at the end of each course, and students receive their grades within two weeks of their examination. Because the number of students enrolled in the Macroeconomics course exceeds the room sizes available at the School of Economics and Management, the students are divided into three classes, taught simultaneously by different teachers.

We conducted the experiment during the Macroeconomics course within the sequence, the second of the four courses. This enables us to use the points in the Microeconomics course to measure performance at baseline. The correlation between Macroeconomics and Microeconomics course points is high: the raw correlation amounts to 0.44; the correlation conditional on baseline student characteristics amounts to 0.87.

As pre-registered, we will use baseline performance as a control variable, allowing us to

[5] Lund University is the fourth largest university in Sweden, with a total enrollment of 40,000-50,000 students.

[6] The points are mapped into grades following this rule: F = 0–49 points, E = 50–54 points, D = 55–64 points, C= 65–74 points, B = 75–84 points, A = 85–100 points.

greatly increase the statistical power of the analysis and to study heterogeneous treatment effects. The Macroeconomics grade was determined by 85% exam performance and 15% performance in two homework assignments.[7] We incentivized students' final letter grade in the Macroeconomics course; however, we pre-registered the total number of points in the course (0-100) as the primary outcome because it captures student learning with more detail.[8]

2.2 Experimental design, procedures, and data collection

Experimental design

The experimental design, timeline, and sample sizes are displayed in Figure 1. The design consists of three treatment groups and one control group. In a baseline survey, we first elicited students' preferred incentive scheme ("goal") out of two possible incentive schemes: Receiving SEK 2,000 ($220 at the time) if their Macroeconomics grade was C or higher or SEK 0 otherwise (the low goal, referred to as "Goal C" to the students), or receiving SEK 4,000 ($440) if their Macroeconomics grade was an A or SEK 0 otherwise (the high goal, referred to as "Goal A").

Two days after the students filled in the baseline survey, we assigned the students randomly to the treatment and control conditions. The random assignment was stratified by baseline grade, the chosen goal, and gender.[9]

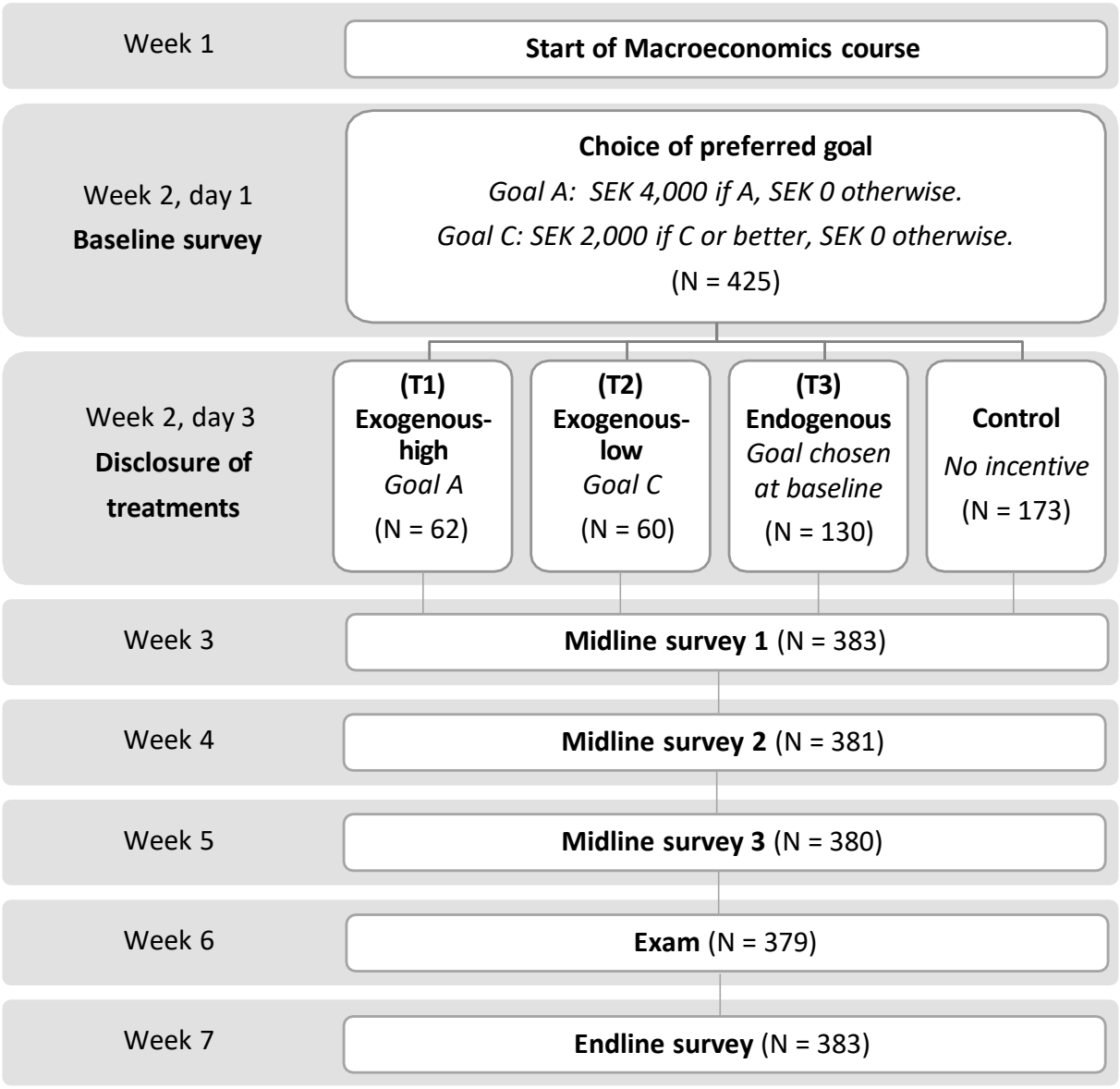
The first treatment group was randomly assigned to Goal A ("exogenous-high treatment," 15% of students), the second treatment group was randomly assigned to Goal C ("exogenous-low treatment," 15% of the students), and the third treatment group was randomly assigned to their preferred goal ("endogenous treatment," 30% of the students). The control group was not offered any incentives (40% of the students). The treatment shares were chosen to maximize the power of the pre-registered hypothesis tests. We explain the procedures in detail next.

[7] One of the homework assignments was completed individually, while the other was a group assignment.

[8] Results are similar qualitatively and quantitatively if we instead use students' exam points or final letter grade (linearly assigning points from F=0 to A=5) in the course.

[9] Accordingly, these are also the main sources of treatment effect heterogeneity that we will study.

Figure 1: Timeline of the experiment, treatment conditions, and sample sizes



Note: The students were randomly divided into the treatment conditions according to the following sample shares: exogenous-high treatment (15% of the sample); exogenous-low treatment (15%); endogenous treatment (30%); control group (40%). The random assignment was stratified by gender, baseline grade (three categories), and the chosen goal at baseline (A- or C-goal).

Recruitment and baseline survey

In total, 425 students participated in the experiment – 193 students in the Fall semester and 232 students in the Spring semester – accounting for 81% of all students who took the Macroeconomics course during these semesters. We recruited students for the experiment during the second lecture week of the Macroeconomics course. One week after the students had started

the course, one of the researchers and a research assistant entered each Macroeconomics class 20 minutes before the end of the lecture to recruit the students and administer a baseline survey. The experiment was not pre-announced, meaning that the students were unaware of the study until the end of this lecture. The students were asked to read and sign a consent form, as well as to fill in a baseline survey during class to participate in the experiment. The paper-based consent form mentioned the broad purpose and title of the study (Academic Interest and Motivation Study) and explained that participants needed to fill in five surveys: one baseline survey to be completed immediately (duration 15 minutes), three midline surveys in each of the three following weeks (duration 30 seconds each), and one endline survey after the Macroeconomics exam (duration 5 minutes). The students were also informed that they would be paid SEK 250 ($28) for completing all three surveys plus some could be offered additional rewards based on their grades.[10]

To our knowledge, all students who attended the lecture agreed to participate in the experiment. This is important as it ensures no self-selection into participating in the study. We also offered extra sessions for students who were not present during the lectures to complete the baseline survey. However, participation in these sessions was limited, and only 26 students entered through the extra sessions. Thus, selection into experiment participation was mainly determined by selection into lecture attendance.[11]

After filling in the consent form, the students entered a link on their laptop or phone browser to fill in the baseline survey during the remaining 15 minutes of class.[12] During the baseline survey, the students provided their social security numbers, enabling us to match their survey responses to their university performance records. Importantly, this allowed us to get the grades of every student who participated in the study, regardless of whether they completed the rest of the surveys. Moreover, we elicited students' baseline characteristics, such as their study habits and socioeconomic background (see Section 3 for details). We asked the participants for their email addresses so that we could contact them for the follow-up (midterm and endline) surveys.

Towards the end of the baseline survey, we asked students to indicate their preferred incen-

[10] The consent form, surveys, and emails the students received are displayed in Appendix 5.
[11] Appendix Table A.8 shows that all results hold, qualitatively and quantitatively, when we drop these students.
[12] We allowed three students to fill in the survey at home instead because they had to leave the lecture early.

tive scheme. The students could choose between receiving SEK 4,000 for reaching course grade A and SEK 0 otherwise (referred to as "Goal A") or receiving SEK 2,000 for getting at least course grade C and SEK 0 otherwise (referred to as "Goal C"). The students were instructed to select the goal carefully. They were informed that there was a 30% chance that they would receive the chosen goal and that with the remaining 70% they would be randomized to a goal or the control group. Eliciting the preferred goal was thus incentive-compatible: students could not benefit from hiding their true preferences. The students could neither opt out of choosing a goal nor propose an alternative goal.[13] Finally, we elicited the students' beliefs about their performance if they would be assigned Goal C, Goal A, or no goal.

At the end of the survey, the students were asked to leave the signed consent form on the table and leave in silence.

Treatment disclosure to the students

All participants received an email with their assigned treatment two days after the baseline survey. Students assigned to the control group were told that their payment would be independent of their grades. Students in the treatment groups were told whether they received Goal C or Goal A. In addition, the students were informed whether they received their goal randomly or because they had chosen it. Appendix Section B.3 displays screenshots of the announcement emails. The number of students is 62 in the exogenous-high treatment (Goal A), 60 in the exogenous-low treatment (Goal C), 130 in the endogenous treatment, and 172 in the control group (see Figure 1).

Data collection on outcomes

In the three weekly midline surveys, students were asked to report their interest in the Macroeconomics course, target grades, study hours, and intention to continue their studies at the School of Economics and Management. Of the 425 participants, 389, 381, and 380 students completed the first, second, and third midline surveys (see Figure 1). The participation rate was very high,

[13] Students were fully informed, however, that they could opt out of the study altogether if they preferred not being assigned the incentives. None of the participants opted out.

94% in the treated groups and 88% in the control group. There were no differences in survey participation rates across the different treatment groups.

The endline survey was sent five days after the students took the final exam but before they knew their grades. The participation rate was again high, 93% on average in the treated groups and 86% in the control group, with no notable differences between the treatment groups. The endline survey contained the same questions as the midline surveys. It also included a broader set of questions capturing the students' study behaviors and attitudes.

After the students received their final grades, we obtained administrative university records containing student performance in the Macroeconomics course (exam participation, points in the exam and homework assignments, and final grade). We merged this information via student identifiers to determine each student's payment.[14] We could match the records for *all* students in the study, regardless of whether they filled out the midline and endline surveys. Students received their final payment within two weeks of obtaining their Macroeconomics grade.[15] Finally, we also obtained grades and points for the Finance and International Economics courses after the students had finished the entire course sequence.

Pre-registration

The study was pre-registered before data collection at the AEA RCT registry (AEARCTR-0003507). We registered the procedures, the share of students in each experimental condition, and the stratifying variables. We also registered the primary outcome (points in the Macroeconomics course) and that we would control for *all* student-level characteristics collected at baseline to increase precision.[16] Power calculations based on a sample size of 400 students

[14]While students can retake the Macroeconomics exam several months later, incentives were paid only based on grades in the first possible exam. This was communicated to the students in advance. Among all students who participated in the experiment, 46 did not take the first exam. Out of those students, 13 completed a retake exam in the same year, 5 completed the course in the following year; 28 did not complete any exam, neither in the same nor in the following year, and thus likely dropped out.

[15]Most students received the payment through Swish, a mobile pay app widely used in Sweden. The rest received it through bank transfer.

[16]We write: "We will use parametric regressions to explain the students' outcome depending on the students' treatment, controlling for all the administrative and survey baseline variables (Microeconomics grade, gender, study time, program, spending, motivation, class, friends, study hours, etc)." Table A.1 displays the full set of controls. Since students' characteristics, and particularly previous grades, are very good predictors of their grades, controlling for these characteristics (which by randomization are orthogonal to the treatments) greatly increases

showed that we have 80% power to detect a treatment effect of ± 0.22 test score standard deviations (two-sided test) when pooling all three treatments. While we expected the interventions to have positive effects, we pre-registered two-sided tests to allow for the possibility of adverse or unintended effects. We also pre-registered heterogeneity analyses, notably heterogeneity by students' ability and gender.

Finally, we registered a large set of secondary outcomes based on survey data. Since the specific way we study the survey data was not pre-registered, we view the analysis of secondary outcomes as exploratory.

3 Data and descriptive statistics

3.1 Pre-assignment data

Baseline characteristics

We collected performance at baseline using administrative university records on Microeconomics course grades. In addition, we elicited students' characteristics during the baseline survey. These characteristics include socioeconomic variables, study habits and motivation, prospective major choices, number and identity of study partners (up to three), and selected behavioral characteristics such as willingness to take risk and patience. We also elicited students' preferred goal and asked them about their target grade for the course and their beliefs (probability of reaching either of the goals under each of the treatments).[17]

Table 1 presents selected summary statistics. The majority of students are male (56%) and intend to major either in Business Administration (39%), Economics (27%), or Political Sciences (20%). Students were generally active in the prior course (Microeconomics). On average, the students attended 80 percent of the lectures in the Microeconomics course and spent about 24 hours per week studying besides attending the lectures. Their final grade in the Microeconomics course was between D and C on average. Most students reached at least a C

our statistical power.

[17] The full list of variables is presented in Table A.1.

(57% of students), and 6% reached an A. The students were eager to improve their grades: at baseline, 15% of them aimed for an A in the Macroeconomics course, and 89% aimed for at least a C. The average target grade was between C and B.

Preferred goal

Toward the end of the survey, the students indicated their preferred goal; 14% of the students chose Goal A and 86% of the students chose Goal C.

The students chose their goals carefully, aligning with their ambition, ability, and willingness to take risk (see Panels A–C of Figure 2). Nearly half of the students who received grade A in the Microeconomics course at baseline chose Goal A (48%). In contrast, only 21% of students who received a B at baseline chose Goal A, and less than 10% of students who received an F–C at baseline chose Goal A. Similarly, 60% of students who had a target grade of A chose Goal A, whereas only 12% of students who had a target grade of B chose the Goal A, and less than 5% of students who had a target grade of E–C chose Goal A. Moreover, out of those students who indicated that their target grade was C, 99% chose Goal C. Finally, we also find that a higher willingness to take risk is strongly associated with choosing Goal A (see Figure 2, Panel C).

Beliefs

After the students chose their goal, we elicited beliefs about their performance under each treatment and the control condition, respectively. Specifically, we asked them about (a) the grade they expected to receive if given no goal, if given Goal A, or if given Goal C, (b) their subjective probability of reaching an A if given Goal A, and (c) their subjective probability of reaching a C if given Goal C.

The students believed that the incentives would increase their performance. On average, the expected grades are lowest when given no goal (slightly lower than C), higher when given Goal C (between C and B), and highest when given Goal A (close to B). No student expected Goal C to backfire (i.e., to lead to a lower grade than no goal), and only seven students anticipated

Table 1: Baseline characteristics: Summary statistics

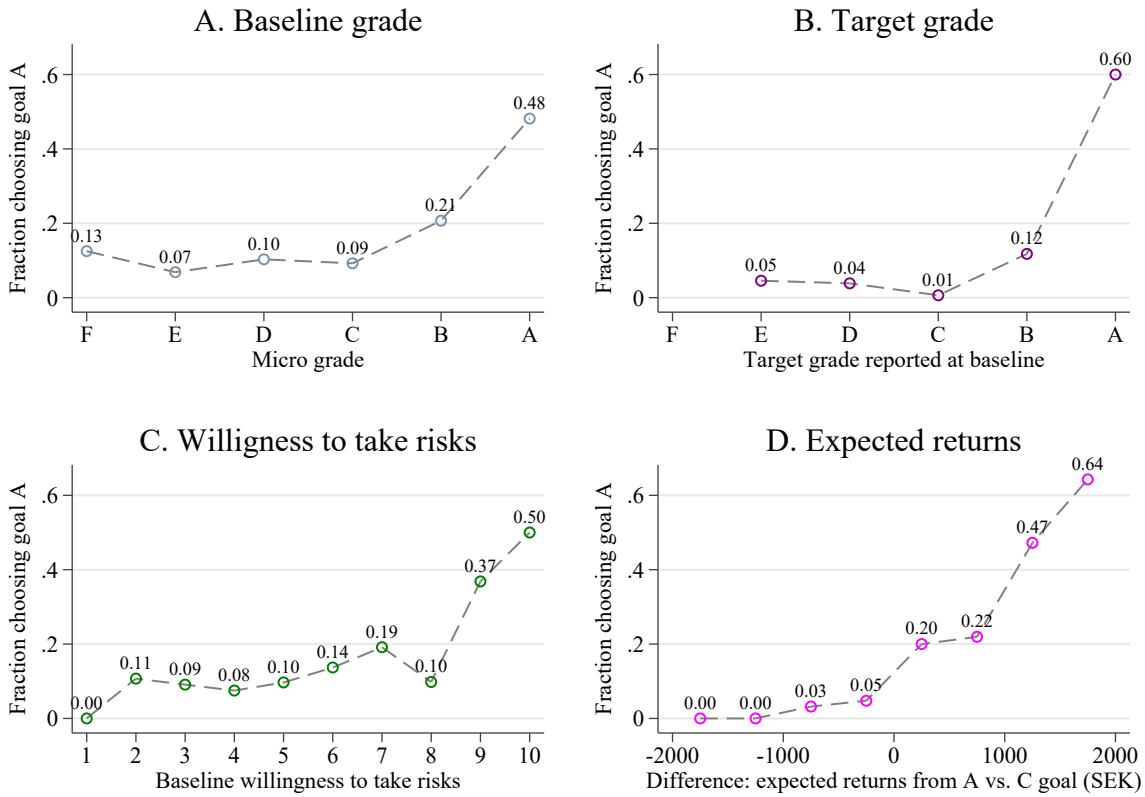
Variable	(1) Mean	(2) SD	(3) Min	(4) Max	(5) Obs.
A. Background					
Female	0.44	-	0	1	425
Monthly spending (1,000 SEK)[(1)]	3.85	2.53	0.2	18.0	425
Willingness to take risks	5.69	2.04	1	10	425
B. Baseline performance					
Micro points (std.)	−0.00	1.00	−2.8	1.8	425
Micro grade A	0.06	0.24	0	1	425
Micro grade at least C	0.57	0.50	0	1	425
Micro: study hours	23.60	11.07	2	60	425
Micro: attendance (%)	79.29	22.60	0	100	425
C. Prospective major choice					
Business Administration	0.39	-	0	1	425
Economics	0.27	-	0	1	425
Political Science	0.20	-	0	1	425
Philosophy & Economics	0.04	-	0	1	425
Accounting	0.04	-	0	1	425
Other	0.07	-	0	1	425
D. Target grade					
Target grade[(2)]	3.50	1.00	1	5	425
Target grade A	0.15	-	0	1	425
Target grade C	0.89	-	0	1	425
E. Beliefs					
Expected grade if no incentives[(2)]	2.87	1.02	0	5	425
Expected grade if assigned A-goal[(2)]	3.92	1.03	0	5	425
Expected grade if assigned C-goal[(2)]	3.49	0.69	0	5	425
Subj. prob. to reach A if given as goal	0.45	0.25	0	1	425
Subj. prob. to reach C if given as goal	0.80	0.20	0	1	425
F. Choice of goal					
A-goal	0.14	0.35	0	1	425
G. Treatment					
(T1) Exogenous-high	0.15	-	0	1	425
(T2) Exogenous-low	0.14	-	0	1	425
(T3) Endogenous	0.31	-	0	1	425
Control group (No incentives)	0.41	-	0	1	425

Note: All variables come from the baseline survey, except for the Microeconomics points and grade (administrative records) and the treatments. The Microeconomics points are standardized (mean zero, standard deviation one).

[(1)] Monthly spending excludes the costs of room, food, and transportation. 1 SEK $\approx$ 0.11 USD.

[(2)] Letter grades are recoded into numbers: 0 = F, 1 = E, 2 = D, 3 = C, 4 = B, 5 = A.

Figure 2: Fraction of students choosing the A-goal by selected baseline characteristics



Note: The figure displays the fraction of students who chose Goal A as their preferred reward scheme on the y-axis and different baseline characteristics in each panel on the x-axis (baseline grade, target grade, willingness to take risk, expected monetary return from the A- versus the C-goal). The grades come from administrative student records. The target grade, willingness to take risk, and questions used to compute expected returns come from the baseline survey. To compute expected returns, we used the answers to the question: "Select the probability that you will reach every one of the Goals if you receive it." We multiplied the probabilities with the respective monetary rewards under each incentive scheme. The x-axis in Panel D displays the difference between the expected return when receiving the A-goal and the expected return when getting the C-goal, pooled in bins of 500 SEK (cutoffs at SEK -1,750, -1,250, -750, -250, 250, 750, 1,250, 1,750). Based on 425 observations.

Goal A to backfire. The students were relatively confident about reaching assigned goals. On average, they believed that they would get an A with a probability of 45% if assigned Goal A and that they would reach a C with a probability of 80% if assigned Goal C. These numbers imply an average expected return of Goal A of SEK 1,784 (≈ $162) and an average expected return of Goal C of SEK 1,593 (≈ $145). Within-student differences between the expected returns of Goal A versus Goal C are positively associated with choosing Goal A (see Panel D

of Figure 2).

3.2 Outcome variables

Our primary outcome variable are points in the Macroeconomics course, which is the sum of points in the exam (up to 85 points) and in two homework assignments (up to 15 points). The exam includes five exercises designed to assess students' understanding of the lecture content. Of the 425 students who participated in the survey, 379 took the exam (89%). Out of 100 possible points in the course, the students received between 8 and 97 points, with a mean of 60 points. For easier interpretation, we standardize the points (with mean zero and standard deviation one). In addition to the total course points, we also study exam and homework points separately. Further, we examine how many students achieved a grade of A or a grade of at least C in the course. In our sample, only 5% of the students reached an A, and 44% of the students reached at least a C (see Panel B of Table 2).

To study the persistent effects of our intervention, we use points in the subsequent courses International Economics and Finance for all students who took the corresponding exams; 83% of the students in our sample took the International Economics, and 68% took the Finance exam (see Panel C of Table 2).

In exploratory analyses, we explore underlying study behaviors and attitudes as correlates of performance. Using questions from the midline and endline surveys, we construct indices for target grades (three survey items), effort (eight items), and motivation (nine items; see Appendix A for details on the survey items in each category). The indices are derived by standardizing the variables (with mean zero and standard deviation one) and computing the unweighted mean of all variables within each category. We then standardize these means so that each index has a mean of zero and a standard deviation of one.[18] Performance in the Macroeconomics course is modestly correlated with our indices of target grades (corr = 0.26, $p < 0.001$), effort (corr = 0.14, $p = 0.006$), and motivation (corr = 0.22, $p < 0.001$).[19]

[18] As an alternative approach, we extract the first principal component from all standardized variables within each category and standardize it to have a mean of zero and a standard deviation of one. The results of our analyses remain unchanged.

[19] The correlation between target grades and performance vanishes, however, when we condition on our pre-

Table 2: Outcomes: Summary statistics

Variable	(1) Mean	(2) SD	(3) Min	(4) Max	(5) Obs.
A. Course performance (Macroeconomics)					
Macro points total (std.)	−0.00	1.00	−3.4	2.2	379
Exam points (std.)	−0.00	1.00	−3.1	2.3	379
Homework points (std.)	0.00	1.00	−4.6	1.3	379
Took exam	0.89	-	0	1	425
Course grade A	0.05	-	0	1	379
Course grade at least C	0.44	-	0	1	379
B. Performance in subsequent courses					
International Economics points (std.)	−0.00	1.00	−3.1	1.8	353
Took exam in International Economics	0.83	-	0	1	425
Finance points (std.)	0.00	1.00	−3.5	1.6	289
Took exam in Finance	0.68	-	0	1	425
C. Survey outcomes (indices)					
Target grades (std.)	0.03	0.98	−2.5	1.5	357
Effort (std.)	0.12	0.91	−2.3	2.9	361
Motivation (std.)	0.01	0.99	−2.8	2.3	361

Note: Descriptive statistics of the main outcome variables analyzed. Course points, course grades, and information on exam participation come from administrative records. Course points in Panel B are missing for students who did not complete the exam in the respective course. The survey outcomes in Panel C are indices summarizing several survey items from the midline and endline surveys that belong to a specific topic (see Appendix A for details on how we constructed the indices and Table A.3 for summary statistics of the underlying survey items). The survey outcomes are coded as missing for students who did not provide answers to any of the respective survey items and for students who did not take the Macroeconomics exam. All outcomes, except for the binary outcomes, are standardized (mean zero, standard deviation one).

3.3 Control variables and balancing checks

We pre-registered all variables collected at baseline as control variables. The full list (see Table A.1) includes socioeconomic variables (gender, spending), study habits, motivation, and satisfaction in the Microeconomics course (course attendance, time spent studying, lecture attendance, interest, and satisfaction with the grade), prospective major choices, number of study partners (up to three), behavioral characteristics (survey questions on willingness to take risk, patience, self-control, optimism, ambition), the chosen goal, the target grade, and students' beliefs about their performance under the different treatment conditions. We also control for the

registered controls.

treatment status of students' peers (number of study partners with incentives).

Table A.2 displays the results of balancing checks. We regress each of the pre-registered controls, in total 36 characteristics, on treatment indicators – in a first regression on an indicator for receiving any incentive, and, in a second regression, on indicators for each treatment arm. Out of 144 coefficients tested, only four are statistically significant at the 5% level and at the 10% level each, which if anything is lower than can be expected by chance. We conclude that the randomization was successful in balancing student background characteristics. Nonetheless, as pre-registered, we control for the full set of variables in all main regressions to increase precision and address any remaining concerns about covariate balance.

4 Results

4.1 Main results: Effects on performance

To examine the effect of incentives we conduct an OLS regression of the form:

$$y_{is} = \alpha_0 + \alpha_1 \textit{incentive}_i + X_i'\gamma + \delta_s + \epsilon_i, \tag{1}$$

where y_{ic} denotes the outcome of student i taking the Macroeconomics course in semester s, $\textit{incentive}_i$ is an indicator variable, taking the value of one if the student was in either of the three treatment groups, and zero if the student was in the control group, $X_i'\gamma$ is a vector of the pre-registered control variables, δ_s is a semester indicator (Spring or Fall semester), and ϵ_i is an error term.

We also investigate the effects of the three different treatment arms separately by performing an OLS regression that includes indicator variables for the three treatments, $\textit{exo\_high}_i$ (exogenous-high treatment), $\textit{exo\_low}_i$ (exogenous-low treatment), and $\textit{endo}_i$ (endogenous treatment), with the students in the control group as the reference category:

$$y_{is} = \beta_0 + \beta_1 \textit{exo\_high}_i + \beta_2 \textit{exo\_low}_i + \beta_3 \textit{endo}_i + X_i'\gamma + \delta_s + \epsilon_i. \tag{2}$$

We begin by assessing whether students understood and responded to the incentives by examining their reported target grades in the three midline surveys.[20] The exogenous-high treatment increased the probability of aiming for an A by 51—54 percentage points relative to the control group ($p < 0.001$, Table A.4, columns 1—3), while the exogenous-low and endogenous treatments raised the probability of targeting at least a C by 8—14 percentage points relative to the control group (p between 0.004 and 0.019, Table A.4, columns 4—6). The smaller effect for C-goals than for A-goals reflects that most students (89%) had targeted at least a C at baseline, whereas few students (15%) had targeted an A at baseline, which left more room for the exogenous-high treatment to influence target grades. In sum, students understood the incentives and adjusted their goals accordingly.

Effects on course performance

Figure 3 shows the raw distribution of points in the Macroeconomics course for students in the control and incentives groups. Contrary to our initial hypothesis, the figure shows that students in the incentives group performed worse overall, with the cumulative distribution shifted to the left compared to the control group. It also shows that students in the incentives group not only achieved lower macro points but also received lower final course grades.[21]

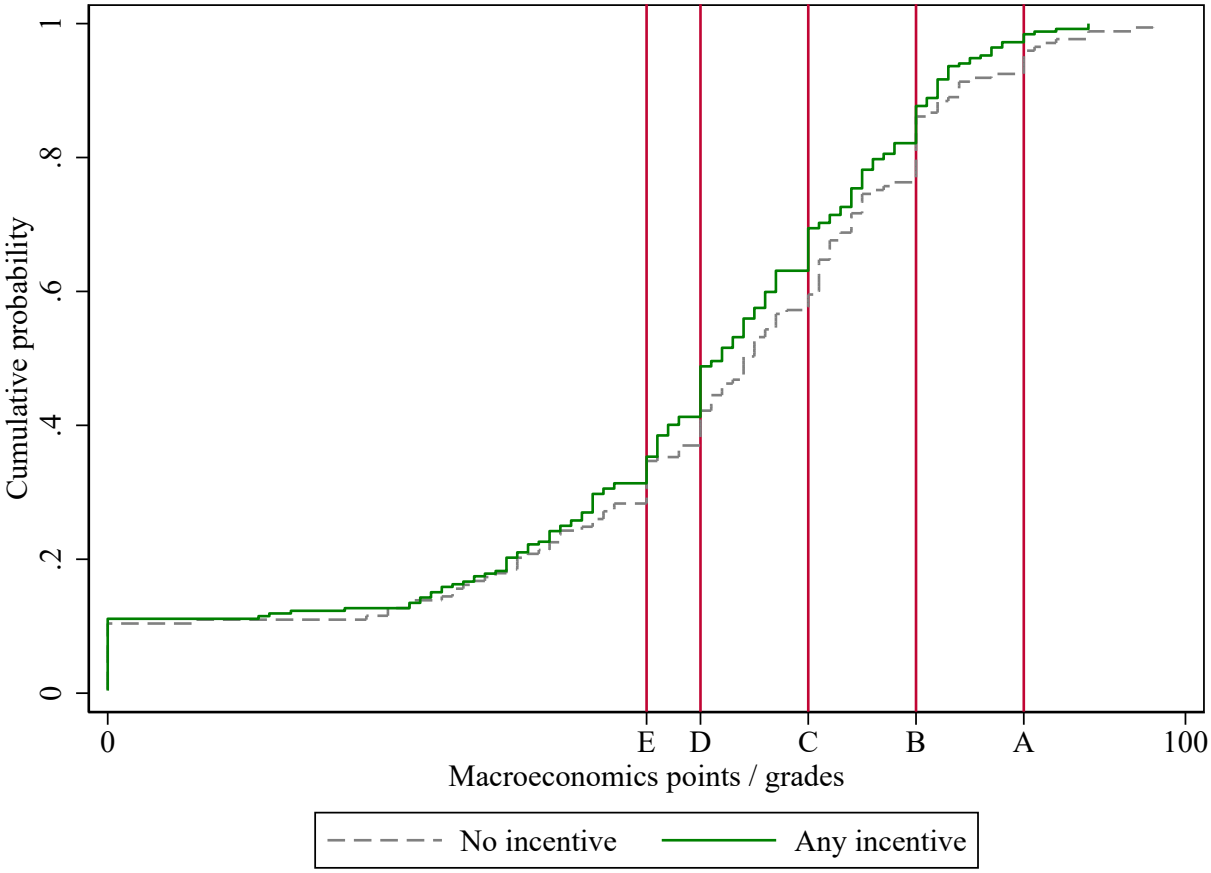
The estimated treatment effects on course performance are displayed in column 1 of Table 3. On average, students who were incentivized performed 0.20 standard deviations lower than control students, a large and statistically significant effect ($p = 0.017$). This negative effect stems primarily from the exogenous-high treatment, which decreases performance by 0.30 standard on average ($p = 0.036$). The effects of the exogenous-low and endogenous treatments are negative as well, but not statistically significant.

Columns 2 and 3 of Table 3 break down the effect into performance on homework assignments (two assignments, 15% of the total score) and the exam (85% of the total score). The total score effect (column 1) is driven entirely by exam performance (column 3), with no impact on

[20] Incentivized students were reminded of the incentives at the start of each survey.

[21] The small discontinuities at each grade threshold reflect an informal policy where teachers review the exams of students who are one point below a higher grade. If possible, they adjust the score by re-evaluating a question, aiming to avoid assigning a final grade that falls just short of the threshold. Note that this policy is long-standing and unrelated to the treatments.

Figure 3: Raw distribution of Macroeconomics points for control vs. incentives groups



Note: The sample consists of all experiment participants (425 students). Students who did not participate in the exam are assigned zero points.

homework performance (column 2). This suggests the negative effects are not due to a general drop in motivation but rather a specific decrease in exam performance. The exam performance drop is statistically significant and sizeable for the exogenous-high and endogenous treatments but insignificant for the exogenous-low treatment (column 3). For homework performance (column 2), we find positive but statistically insignificant point estimates for the exogenous-high and endogenous treatments. Since homework accounts for only 15% of the total score, the treatment effects on the total score closely mirror those on the exam score.

Additional outcomes: course grades and persistence

Although the treatments had adverse effects on average, they might theoretically have helped students attain the incentivized performance thresholds (grade A or grade of at least C). In line

Table 3: Effects of incentives on course performance

	(1)	(2)	(3)
	Dep.var.: Course performance (std.)		
	total	homework	exam
Panel A. Effect of any incentive			
Any treatment (pooled)	–0.20**	0.07	–0.22**
	(0.09)	(0.10)	(0.09)
R-squared	0.47	0.27	0.46
Panel B. Effects of the different incentives			
(T1) Exogenous-high	–0.30**	0.15	–0.34**
	(0.14)	(0.14)	(0.15)
(T2) Exogenous-low	–0.17	–0.07	–0.15
	(0.12)	(0.18)	(0.13)
(T3) Endogenous	–0.17*	0.10	–0.20**
	(0.10)	(0.11)	(0.10)
R-squared	0.47	0.27	0.46
Control mean	0.08	–0.01	0.09
Observations	379	379	379
Pre-registered controls	✓	✓	✓

Note: The table shows results of OLS regressions of course performance outcomes on a treatment indicator for receiving any incentive (Panel A, see equation 1) and on indicators for the different treatment arms (Panel B, see equation 2). Table A.1 displays the full set of pre-registered controls. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

with our headline results, however, the treatments do not help the students reach grades A or C (see columns 1 and 2 in Appendix Table A.5).

We also study the persistence of the negative treatment effects using performance in subsequent courses. We find that the negative effect is confined to the incentivized course grades and does not translate to later courses (see columns 3 and 4 in Appendix Table A.5).

Robustness and validity checks

Several robustness checks support our main findings. The results are qualitatively and quantitatively similar when including fewer control variables (Appendix Table A.6, columns 1–2);

with fewer controls, however, the effect is more noisy and somewhat attenuated compared to our pre-registered specification.

Further, we address concerns about selection into the exam in Appendix Table A.7. Incentives do not affect selection into exam-taking (column 1); moreover, reweighting the main estimates by the inverse of students' predicted probability to take the exam (Wooldridge, 2007) does not affect the results either (column 2).

We also address concerns related to potential self-selection into the experiment through the separate (extra) recruitment sessions. Excluding these students from the analysis does not affect our findings (see Appendix Table A.8), which is expected given that only 26 students were recruited this way, with just 18 of them ultimately completing the exam, representing less than 5% of our total sample.

We also document that spillover effects between treated and control students are unlikely to drive our results; the main concern would be that positive reactions of control students to their peers' treatments could bias our results downwards. First, controlling for the number of peers with incentives in our regressions does not affect the results (Appendix Table A.6, compare columns 3 and 4), and peers' treatment status does not have any statistically significant effect on performance (Appendix Table A.9). Similarly, conducting regressions separately in the subsamples of students who interact more versus less with others shows very consistent patterns across the two groups (Appendix Table A.10). Finally, we document that course performance in the cohorts under study was lower overall than course performance in previous cohorts (Appendix Figure A.1). This pattern is consistent with overall negative treatment effects in the cohorts under study, and does not suggest that the treatments positively affected the control group.[22]

[22] Additionally, we can directly assess whether students in the control group reportedly changed their study behaviors in response to the study. To this end, we draw upon students' answers to an open-ended question at the end of the survey. Only 14% of the students in the control group but 38% of the treated students wrote that the study affected their studying patterns, suggesting that the treatments mainly affected behaviors among the treated students.

4.2 Effect heterogeneity and goal mismatch

Effect heterogeneity by ability and gender

Following our pre-registration and stratification strategy, we focus on baseline achievement (ability) and gender as the main sources of treatment effect heterogeneity. Ability is particularly important in our setup, as first-year students typically enter university with varying levels of preparedness, potentially affecting their response to incentives that are (mis-)aligned with their initial achievement. We proxy ability using performance in the previous course (see Section 3) and classify students as "high-ability" if they earned an A or B in Microeconomics and "low-ability" otherwise.

Table 4, columns 1–2, show that the treatment effects are driven by low-ability students. The exogenous-high treatment reduces low-ability students' performance by 0.38 percentage points (p = 0.033) but does not impact high-ability students, for whom the point estimate is positive but statistically insignificant. The exogenous-low and endogenous treatments show negative but statistically insignificant effects for both groups.

Gender is another potential important source of heterogeneity in how students respond to incentives. For example, men and women may differ in how they value financial rewards or in their levels of confidence (e.g., Angrist and Lavy 2009; Angrist et al. 2014b; Gneezy et al. 2017; Levitt et al. 2016; Campos-Mercade and Wengström 2020). Columns 3 and 4 of Table 4 show treatment effects by gender. The negative effects are driven primarily by male students, whose performance decreases by 0.28 standard deviations in response to any incentive (p = 0.029). The exogenous-high treatment reduces male students' performance by 0.41 standard deviations (p = 0.036) but has no statistically significant effect on female students. This gender difference is unlikely to reflect ability differences, because male and female students perform similarly at baseline (difference of 0.001 standard deviations, p = 0.989). The point estimates for the exogenous-low and endogenous treatments are negative but statistically insignificant for male students, and close to zero for female students.

Table 4: Effect heterogeneity by ability and gender

	(1)	(2)	(3)	(4)
	Dep.var.: Course performance (std.)			
Heterogeneity:	By ability		By gender	
	low	high	males	females
Panel A. Effect of any incentive				
Any treatment (pooled)	–0.21*	–0.13	–0.28**	–0.01
	(0.11)	(0.16)	(0.13)	(0.13)
R-squared	0.42	0.59	0.54	0.54
Panel B. Effects of the different incentives				
(T1) Exogenous-high	–0.38**	0.08	–0.41**	–0.08
	(0.18)	(0.29)	(0.19)	(0.22)
(T2) Exogenous-low	–0.14	–0.15	–0.31	–0.02
	(0.15)	(0.25)	(0.19)	(0.22)
(T3) Endogenous	–0.16	–0.21	–0.20	0.02
	(0.13)	(0.17)	(0.15)	(0.16)
R-squared	0.43	0.60	0.54	0.54
Control mean	–0.18	0.72	0.06	0.10
Observations	267	112	208	171
Pre-registered controls	✓	✓	✓	✓

Note: The table shows OLS regression results of course performance on treatment indicators. The sample is split by ability (high: baseline grade A or B, low: baseline lower than B) and gender (self-reported in the baseline survey). In all regressions, we control for all pre-registered controls. Clustered standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

The role of goal mismatch

Given the heterogeneous treatment effects by ability that we observe, a natural question is whether the effectiveness of incentives decreases when assigned goals are misaligned with students' individual capabilities or aspirations. A core feature of our design is that it allows us to study the effects of goal mismatch.

We define a student as "matched" if the student is a high-ability student assigned the exogenous-high goal or a low-ability student assigned the exogenous-low goal and as "mismatched" if the student is a high-ability student assigned the exogenous-low goal and a low-ability student assigned the exogenous-high goal. We exclude students in the endogenous treatment group from

the analysis, because they could be matched or mismatched depending on their choice of goal at baseline. We conduct an OLS regression of course performance on indicator variables for receiving a matched treatment, $matched_i$, and for receiving a mismatched treatment $mismatched_i$, with students in the control group in the reference category:

$$y_{ic} = \beta_0 + \beta_1 matched_i + \beta_2 mismatched_i + X_i'\gamma + \delta_c + \epsilon_i. \tag{3}$$

Table 5 presents the results of the mismatch analysis. A mismatched treatment decreases performance by 0.32 standard deviations on average (p = 0.029, see column 1 of Table 5). A matched treatment also negatively affects the students, but the effect is smaller and statistically insignificant. The mismatch effect is mostly driven by low-ability students (column 2). For the high-ability students, the mismatch effect is sizeable albeit statistically insignificant (point estimate of -0.20, p = 0.461, see column 3).

Table 5: Effect of goal mismatch

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep.var.: Course performance (std.)					
	Mismatch by ability			Mismatch by chosen goal		
Mismatched	–0.32**	–0.33*	–0.20	–0.36**	–0.39**	0.20
	(0.15)	(0.19)	(0.26)	(0.14)	(0.15)	(0.38)
Matched	–0.10	–0.12	–0.11	–0.15*	–0.18*	0.17
	(0.13)	(0.17)	(0.32)	(0.09)	(0.09)	(0.33)
Sample	all	ability low	ability high	all	chose C	chose A
Observations	269	190	79	379	331	48
R-squared	0.51	0.46	0.79	0.48	0.48	0.36
Control mean	0.08	–0.18	0.68	0.08	–0.00	0.53
Pre-registered controls	✓	✓	✓	✓	✓	✓

Note: The table shows OLS regression results of the treatment effect of matched and mismatched treatments. In columns 1–3, we define a student as "matched" if the student is either a high-ability student and was assigned the exogenous-high goal or a low-ability student and was assigned the exogenous-low goal and as "mismatched" otherwise (see equation 3 for details). In columns 4–6, we define a student as "matched" if they chose the A-goal at baseline and received the exogenous-high or endogenous treatment, or if they chose the C-goal at baseline and received the exogenous-high or endogenous treatment, and as "mismatched" otherwise. In columns 1–3, we exclude students in the endogenous treatment group. In all regressions, we control for all pre-registered controls. Clustered standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

In Table 5, we also quantify mismatch effects by a student's chosen goal at baseline. We define a student as matched if they were assigned the goal they chose at baseline and as mismatched if they were assigned a different goal. In this setup, whether a student is matched or mismatched is random, even for the endogenous treatment group, which always receives a matched treatment.[23] The results on mismatch by chosen goal closely mirror the results on mismatch by ability. The negative mismatch effect stems from students who chose the C-goal at baseline (column 5); the estimates for the sample of students who chose the A-goal are noisy as this sample is small (column 4). We also investigate the mismatch by a student's target grade reported at baseline (see Appendix Table A.11) and find similar results.

We note that both definitions of goal alignment, based on ability and based on goal choice, are proxies rather than perfect measures. Students of similar ability may still differ in what goal best motivates them, depending on their confidence, preferences, or other characteristics. Likewise, when using stated goals as a proxy, we restrict students to choosing between only two options (Goal A or Goal C), and neither of them may reflect their true optimal goal. As a result, even students classified as "matched" in our analysis may experience some degree of misalignment, which could help explain why we find negative, though statistically insignificant, effects even within the matched group.

4.3 Discussion: Why did the incentives backfire?

The negative effects of goal mismatch suggest that adverse incentive effects arise when the incentives are misaligned with a student's originally targeted grade. Supporting this interpretation, we find that the groups that most strongly adjust their target grades in response to incentives also experience the largest performance declines. In Table 6, we present the treatment effects on students' reported target grades (desired course grades). The target grade is measured here by averaging the reported target grades from the three midline surveys. The incentive treatments

[23] To see that, first, consider a student who chose Goal A. This student is mismatched with a probability of 15% (probability to receive the exogenous-low treatment); the student is matched with a probability of 45% (exogenous-high: 15%, endogenous: 30%) and is in the control group with a probability of 40%. Similarly, a student who chose Goal C is mismatched with a probability of 15% (probability of receiving the exogenous-high goal), matched with a probability of 45% (exogenous-low: 15%, endogenous: 30%), and in the control group with a probability of 40%.

lead to strong upward shifts in target grades: the effect amounts to 0.29 standard deviations on average ($p < 0.001$, see column 1 of Table 6).[24] It is primarily driven by the exogenous-high treatment, which induces students to increase their target grades by 0.82 standard deviations on average ($p < 0.001$). Furthermore, the effect is mainly driven by male and low-ability students (see columns 2–5 of Table 6). This is noteworthy as these groups—the exogenous-high treatment group, males, and low-ability students—are also the ones experiencing the largest negative treatment effects on course performance (as previously displayed in Table 4), reinforcing the link between shifts in target grades and detrimental performance effects.

One potential mechanism is that incentives distorted students' internal reference points. Initially, students may have their own ambitious but attainable target grade that helps them stay motivated and perform well (Heath et al., 1999, Koch and Nafziger, 2016, 2020, Clark et al., 2020, González-Jiménez, 2024). However, incentives to reach a performance target can potentially shift this reference point to a less optimal one from a performance perspective. For example, low-ability students with a high goal may realize while writing their exam that they will not reach their new target grade, prompting negative thoughts (e.g., Curci et al., 2013) and lack of motivation.

Prospect theory could provide a theoretical underpinning for why shifts in reference points prove detrimental. Under diminishing sensitivity, a key feature of prospect theory, the marginal rewards from increased effort diminish as the reference points become distant and unattainable. Similarly, each additional shortfall may feel less and less consequential. These arguments align with earlier evidence that difficult-to-reach performance goals can produce negative affect, stress, and lowered effort, especially if individuals perceive failure as imminent (Höpfner and Keith, 2021, Campos-Mercade et al., 2022). [25]

An alternative explanation as to why incentives backfired is through crowding out of in-

[24] This is in line with the findings from Table A.4 where the effect on target grades C or A in each midline survey was assessed separately.

[25] Some earlier laboratory studies find that higher reference points can increase effort, possibly due to loss aversion (e.g., Abeler et al. 2011, Campos-Mercade et al. 2024a). However, these studies differ in that individuals can directly control their output (e.g., number of tasks solved), which may activate different motivational mechanisms than when performance is partly stochastic. Indeed, previous research shows that task-based goals tend to be more effective than performance-based goals for students (Fryer 2011, Clark et al. 2020; see also Koch and Nafziger 2020 on the distinction between narrow and broad goals).

Table 6: Treatment effects on target grades

	(1)	(2)	(3)	(4)	(5)
	Dep.var.: Target grades (std.)				
Heterogeneity:	all students	By ability		By gender	
		low	high	males	females
Panel A. Effect of any incentive					
Any treatment (pooled)	0.29***	0.34***	0.09	0.50***	0.11
	(0.08)	(0.10)	(0.14)	(0.12)	(0.12)
R-squared	0.60	0.58	0.73	0.65	0.65
Panel B. Effects of the different incentives					
(T1) Exogenous-high	0.82***	0.90***	0.49***	1.09***	0.62***
	(0.14)	(0.20)	(0.16)	(0.20)	(0.22)
(T2) Exogenous-low	0.10	0.20*	–0.32	0.36**	–0.17
	(0.09)	(0.11)	(0.25)	(0.14)	(0.17)
(T3) Endogenous	0.14*	0.14	0.08	0.26**	0.02
	(0.08)	(0.10)	(0.16)	(0.12)	(0.13)
R-squared	0.64	0.63	0.77	0.70	0.70
Control mean	0.49	–0.42	0.71	–0.19	–0.20
Observations	357	249	108	191	166
Pre-registered controls	✓	✓	✓	✓	✓

Note: The table shows results of OLS regressions of target grades on treatment indicators. Target grades are the average of the target grades reported during the three midline surveys; the average is then standardized (mean zero and standard deviation one). The sample is split by ability (baseline grade A or B versus baseline grade C or lower, columns 2 and 3) and gender (reported in the baseline survey, columns 4 and 5). In all regressions, we control for all pre-registered controls. Clustered standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

trinsic motivation (which has been observed in other contexts, see e.g., Deci and Ryan 1985; Gneezy and Rustichini 2000; Bénabou and Tirole 2003; Prendergast 2008). However, several patterns in our data speak against this mechanism. Examining students' self-reported motivation throughout the course, we do not find any meaningful negative effects (see column 1 of Table A.12). If anything, the endogenous treatment raises motivation slightly (coefficient of 0.17 standard deviations, $p = 0.050$). Neither matched nor mismatched goals affect motivation negatively (see Figure A.2). In addition, performance in homework assignments did not deteriorate, suggesting no general decline in motivation for the course (see column 5 of Table 3).

Moreover, contrary to the crowd-out hypothesis, our effort variable based on self-reported

study behaviors does not decline in response to the treatments (see column 2 of Table A.12). In fact, we find positive albeit short-lived effects on effort as the course progresses: one week after the start of the intervention, effort was significantly higher in the incentive conditions than in the control group (effect of 2.38 hours per week, $p = 0.012$, see column 2 of Table A.13); however, the effect vanishes toward the end of the lecture period as students in both control and treatment groups increase their study hours to nearly 30 hours per week on average (Table A.13, column 3). Importantly, increases in target grades and reported effort are most pronounced among male and low-ability students (see Fig A.3), precisely the groups whose performance react most negatively to incentives. This heterogeneous pattern further contradicts a general crowding-out of motivation.

The absence of measurable effects on effort and performance during the course suggests that the negative impact may have emerged later—specifically during the final, typically intensive, study period before the exam. At this point, students may have realized that their reference point was out of reach, leading to demotivation in the final stretch of the course. Students with high reference points might also have experienced choking under pressure, either in the lead-up to or during the exam itself (e.g., Baumeister and Showers, 1986, Dohmen and Falk, 2011, Ariely et al., 2009). This pressure could have impaired attention and concentration both during exam preparation and on the exam day.[26]

Overall, our evidence indicates that the negative impact of external incentives on student performance relates specifically to shifts in students' goals, which may disrupt their usual motivational targets by setting new reference points. Importantly, we do not find evidence for a general crowding-out of intrinsic motivation. Rather, the reduction in performance appears linked to reference-dependent mechanisms operating through two possible channels: first, students who perceive their goals as unattainable might experience diminished motivation and reduced value of effort during the crucial later stages of the course or even on the exam itself; second, students facing excessively challenging goals may choke under pressure, further impairing performance. While our current data cannot conclusively disentangle these two mechanisms, our

[26] Another potential channel for low exam performance could be related to exam-taking strategies. Students might have adopted a riskier approach in hopes of securing a high grade. However, since the exam did not include multiple-choice questions, there was no obvious opportunity to pursue such a strategy.

findings support the idea that the observed negative effects are related to incentives altering students' target grades.

5 Conclusion

Incentives to achieve performance goals are prevalent across many economic domains. Yet their effectiveness remains debated and controlled evidence from field settings is scarce. This paper employs a field experiment in an educational setting to examine how students respond to incentives for achieving performance goals of varying difficulty. We find that performance-based incentives generally reduce rather than enhance student performance, with treated students performing worse than those in the control group. These negative effects are particularly pronounced among male students and low-ability students assigned high performance targets. Allowing students to set their own goals partially, but not completely, mitigates these adverse effects. Using rich survey data collected before, during, and after the course, we identify shifts in students' target grades and discuss goal misalignment as the likely mechanism behind the adverse incentive effects.

These findings add to our understanding of the effects of incentivized performance goals on performance. Our findings challenge traditional economic theory, which predicts that incentives enhance performance. Specifically, our results highlight potential pitfalls arising when performance goals alter individuals' targets and fail to align with their personal motivations and abilities. These results are important given the prevalence of threshold-based incentive schemes in many institutional settings, including business, healthcare, and education. Policymakers and educators should exercise caution in implementing performance-based incentives, as these incentives can influence individuals' own goals and reference points, negatively impacting overall performance. When individuals' abilities and personal targets cannot be precisely identified or controlled, alternative policies such as non-incentivized interventions or policies that better align with individual goals may be more effective in fostering positive outcomes.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.

Angrist, J. and Lavy, V. (2009). The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized. *American Economic Review*, 99(4):1384–1414.

Angrist, J., Oreopoulos, P., and Williams, T. (2014a). When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources*, 49(3):572–610.

Angrist, J., Oreopoulos, P., and Williams, T. (2014b). When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards. *Journal of Human Resources*, 49(3):572–610.

Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2):451–469.

Baumeister, R. F. and Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology*, 16(4):361–383.

Becker, G. S. (1967). *Human capital and the personal distribution of income: An analytical approach*. Ann Arbor: University Michigan Press.

Behrman, J. R., Parker, S. W., Todd, P. E., and Wolpin, K. I. (2015). Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools. *Journal of Political Economy*, 123(2):325–364.

Bellés-Obrero, C. (2020). Who is learning? A field experiment comparing three different incentive schemes in the same educational setting. Technical report, Working Paper.

Bénabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520.

Bettinger, E. (2012). Paying to Learn: The Effect on Financial Incentives on Elementary School Test Scores. *Review of Economics and Statistics*, 94(3):686–698.

Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.

Campos-Mercade, P., Currarini, S., Marchiori, C., and Mengel, F. (2022). Performance thresholds, effort and risk-taking. *SSRN*.

Campos-Mercade, P., Goette, L., Graeber, T. W., Kellogg, A., and Sprenger, C. (2024a). *" De Gustibus" and Disputes about Reference Dependence*. Harvard Business School.

Campos-Mercade, P., Meier, A. N., Meier, S., Pope, D. G., Schneider, F. H., and Wengström, E. (2024b). Incentives to vaccinate. Technical report, National Bureau of Economic Research.

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., and Wengström, E. (2021). Monetary incentives increase covid-19 vaccinations. *Science*, 374(6569):879–882.

Campos-Mercade, P. and Wengström, E. (2020). Threshold incentives and academic performance. *Mimeo, Lund University*.

Casey, M. D., Cline, J., Ost, B. E. N., and Qureshi, J. A. (2018). Academic probation, student performance, and strategic course-taking. *Economic Inquiry*, 56(3):1646–1677.

Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.

Clark, D., Gill, D., Prowse, V., and Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4):648–663.

Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*, 61(12):2926–2944.

Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2018). Goal setting in the principal–agent model: Weak incentives for strong performance. *Games and Economic Behavior*, 109:311–326.

Cornwell, C. M., Lee, K. H., and Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, 40(4):895–917.

Curci, A., Lanciano, T., Soleti, E., and Rimé, B. (2013). Negative emotional experiences arouse rumination and affect working memory capacity. *Emotion*, 13(5):867.

De Meza, D. and Webb, D. C. (2007). Incentive design under loss aversion. *Journal of the European Economic Association*, 5(1):66–92.

De Paola, M., Scoppa, V., and Nisticò, R. (2012). Monetary incentives and student achievement in a depressed labor market: results from a randomized experiment. *Journal of Human Capital*, 6(1):56–85.

Deci, E. and Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6):627.

DellaVigna, S. and Pope, D. (2017). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

Dittmann, I., Maug, E., and Spalt, O. (2010). Sticks or carrots? optimal ceo compensation when managers are loss averse. *The Journal of Finance*, 65(6):2015–2050.

Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–90.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2024). The effect of incentives in nonroutine analytical team tasks. *Journal of Political Economy*, 132(8):2695–2747.

Friebel, G., Heinz, M., Krueger, M., and Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, 107(8):2168–2203.

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4):1755–1798.

Gneezy, U., List, J. A., Livingston, J. A., Sadoff, S., Qin, X., and Xu, Y. (2017). Measuring success in education: the role of effort on the test itself. Technical report, National Bureau of Economic Research.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Gneezy, U. and Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, 29(1):1–17.

Goerg, S. J. and Kube, S. (2012). Goals (th)at work–goals, monetary incentives, and workers' performance. *MPI Collective Goods Preprint*, 2012/19.

Goerg, S. J., Kube, S., and Radbruch, J. (2019). The effectiveness of incentive schemes in the presence of implicit effort costs. *Management Science*, 65(9):4063–4078.

Golman, R. and Loewenstein, G. (2012). Expectations and aspirations: Explaining ambitious goal-setting and nonconvex preferences. Technical report, mimeo.

Gomez-Minambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology*, 33(6):1223–1239.

González-Jiménez, V. (2024). Incentive design for reference-dependent preferences. *Journal of Economic Behavior & Organization*, 221:493–518.

González-Jiménez, V., Dalton, P. S., and Noussair, C. N. (2019). Bonuses and loss aversion. *Working Paper*.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive psychology*, 38(1):79–109.

Herranz-Zarzoso, N. and Sabater-Grande, G. (2018). Monetary incentives and self-chosen goals in academic performance: An experimental study. *International Review of Economics Education*, 27:34–44.

Herweg, F., Müller, D., and Weinschenk, P. (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, 100(5):2451–2477.

Höpfner, J. and Keith, N. (2021). Goal missed, self hit: Goal-setting, goal-failure, and their affective, motivational, and behavioral consequences. *Frontiers in Psychology*, 12:704790.

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2):601–626.

Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1):212–227.

Koch, A. K. and Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, 162:305–351.

Koch, A. K. and Nafziger, J. (2020). Motivational goal bracketing: An experiment. *Journal of Economic Theory*, 185:104949.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lacetera, N., Macis, M., and Slonim, R. (2013). Economic rewards to motivate blood donations. *Science*, 340(6135):927–928.

Lazear, E. P. (2000a). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

Lazear, E. P. (2000b). The power of incentives. *American Economic Review*, 90(2):410–414.

Leuven, E., Oosterbeek, H., and Van der Klaauw, B. (2010). The effect of financial rewards on students'achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

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

Levitt, S., List, J. A., and Sadoff, S. (2017). The effect of performance-based incentives on educational achievement: Evidence from a randomized experiment. *Working Paper*.

Montalbán, J. (2023). Countering moral hazard in higher education: The role of performance incentives in need-based grants. *The Economic Journal*, 133(649):355–389.

Ockenfels, A., Sliwka, D., and Werner, P. (2014). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

O'Donoghue, T. and Sprenger, C. (2018). Reference-dependent preferences. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 1–77. Elsevier.

Prendergast, C. (2008). Intrinsic motivation and incentives. *American Economic Review*, 98(2):201–205.

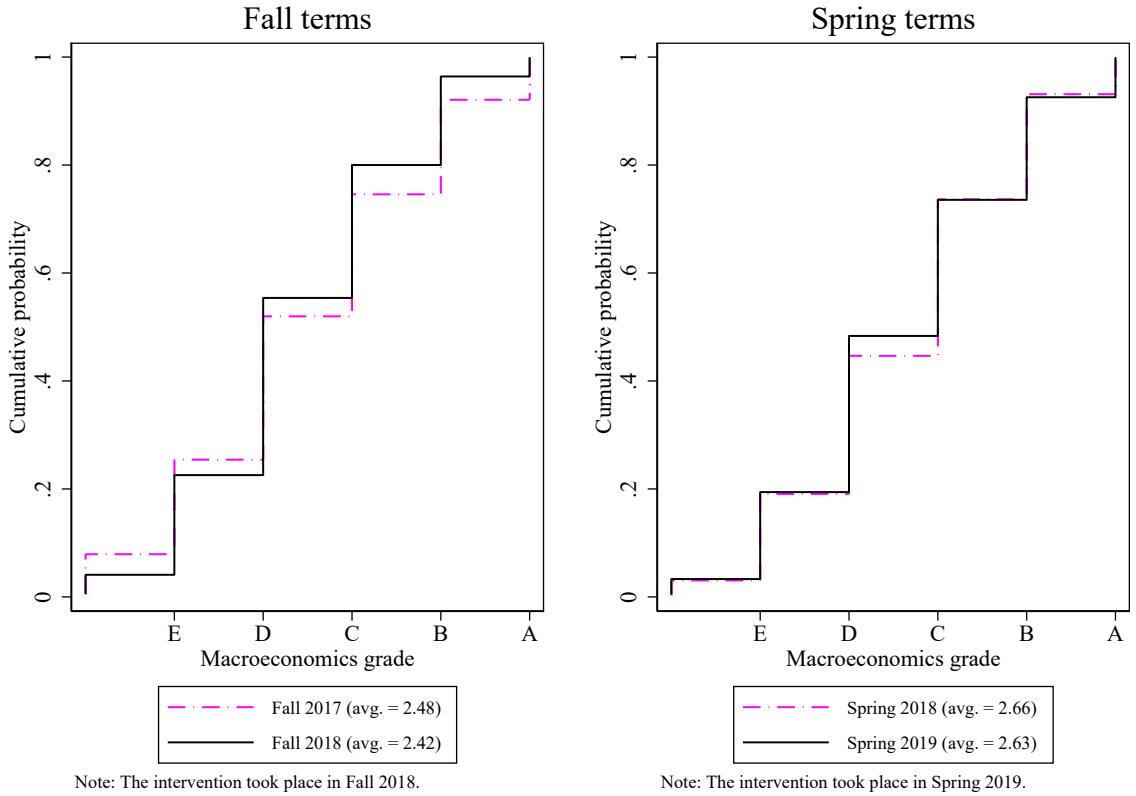
Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3):614–646.

Woerner, A., Romagnoli, G., Probst, B. M., Bartmann, N., Cloughesy, J. N., and Lindemans, J. W. (2024). Should individuals choose their own incentives? evidence from a mindfulness meditation intervention. *Management Science*.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2):1281–1301.

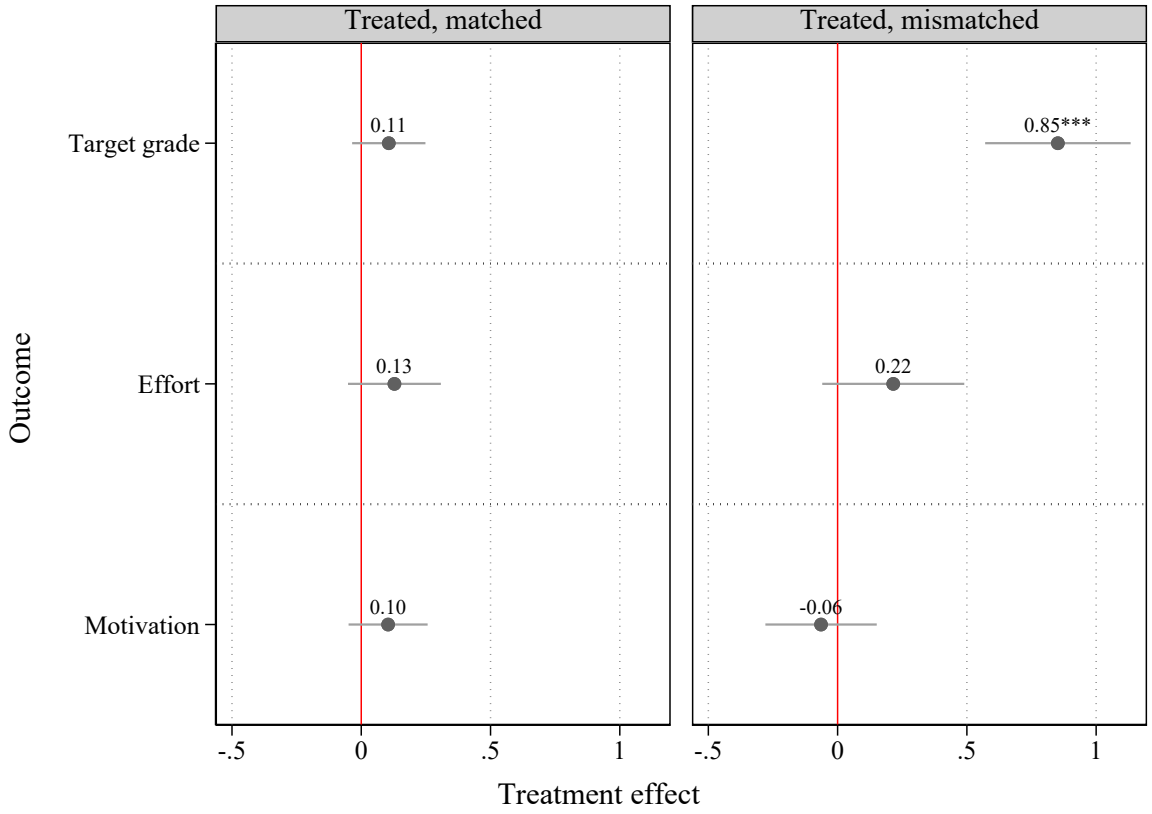
Appendix

Figure A.1: Comparison of course performance with previous cohorts



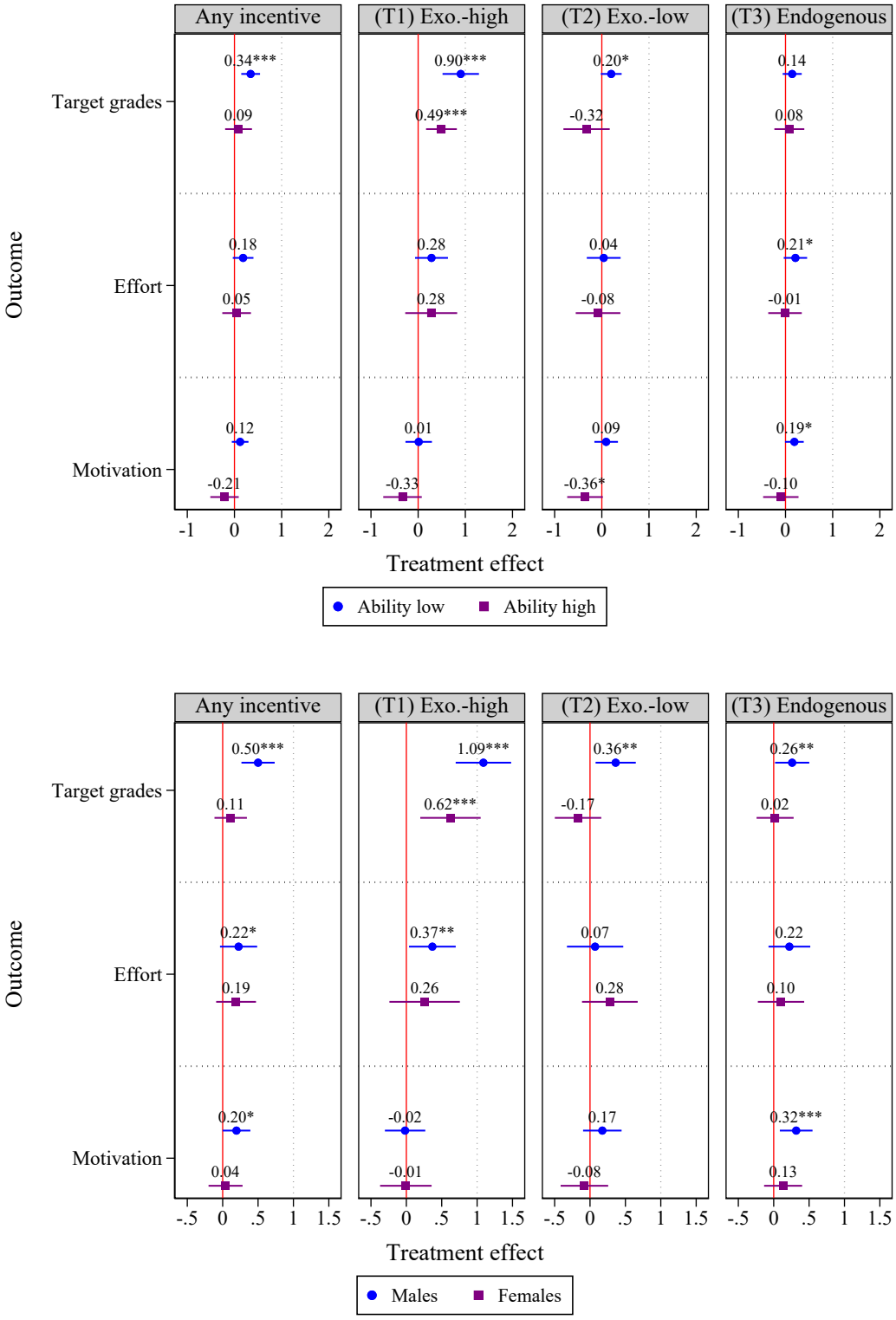
Note: The figure displays CDFs of course grades for all students who completed the macroeconomics course in the Fall 2017–Spring 2019 semesters and took the first (main) exam. Students who only took the retake exam are excluded. To compute averages, the letter scale is transformed into numbers (F = 0, E = 1, ..., A = 5). The data comes from administrative records provided by Lund University (N = 875).

Figure A.2: Effects of matched and mismatched treatments on survey outcomes



Note: The figure displays results of OLS regressions of survey outcomes on indicators for matched and mismatched treatments, based on regression equation 3. The left panel presents the effects of the matched treatments, and the right panel the effects of the mismatched treatments. All regressions include the pre-registered controls. The whiskers present 90%-confidence intervals based on robust standard errors. *p<0.1, **p<0.05, ***p<0.01.

Figure A.3: Effect heterogeneity by ability and gender



Note: All regressions include the pre-registered controls. The whiskers present 90%-confidence intervals based on robust standard errors. *p<0.1, **p<0.05, ***p<0.01.

Table A.1: Descriptive statistics of control variables

Variable	(1) Mean	(2) SD	(3) Min	(4) Max
Stratifying variables				
Choice: goal A	0.14	-	0	1
Baseline performance low	0.27	-	0	1
Baseline performance middle	0.47	-	0	1
Baseline performance high	0.27	-	0	1
Female	0.44	-	0	1
Baseline performance				
Micro: course points (std.)	−0.00	1.00	−2.78	1.80
Cohort and teacher				
Fall cohort	0.45	-	0	1
Spring cohort	0.55	-	0	1
Teacher 1	0.25	-	0	1
Teacher 2	0.60	-	0	1
Teacher 3	0.15	-	0	1
Controls from the survey				
Monthly spending (1,000 SEK)	3.85	2.53	0.20	18.00
Micro: attendance	79.29	22.60	0	100
Micro: study with others (%)	42.83	29.49	0	100
Micro: study hours	23.60	11.07	2	60
Micro: found interesting	3.03	0.64	1	4
Micro: satisfied with grade	2.62	0.92	1	4
Prospective major: Economics	0.27	-	0	1
Prospective major: Business Administration	0.39	-	0	1
Prospective major: Accounting	0.04	-	0	1
Prospective major: Philosophy & Economics	0.04	-	0	1
Prospective major: Political Sciences	0.20	-	0	1
Prospective major: Other	0.07	-	0	1
Prob.: continue at School of Econ & Mgmt	44.88	30.14	0	100
Willingness to take risk	5.69	2.04	1	10
Patience	7.17	1.80	1	10
Self control: resist temptation	3.24	0.96	1	5
Self control: am lazy	2.54	1.03	1	5
Self control: do bad things if fun	3.15	1.07	1	5
Self control: iron self-discipline	3.11	1.01	1	5
Self control: long-term orientation	3.81	0.86	1	5
Self control: want to study more	3.69	1.14	1	5
Optimism: expect the best	2.92	1.15	1	5
Pessimism: can go wrong	2.25	0.96	1	5
Optimism: future	3.52	1.05	1	5
Success: importance	4.09	0.94	1	5
Afraid of failure	3.43	1.20	1	5
High grade will help	3.46	1.09	1	5
Expected grade if no incentives	2.87	1.02	0	5
Expected grade if assigned A-goal	3.92	1.03	0	5
Expected grade if assigned C-goal	3.49	0.69	0	5
Prob. to reach A if given as goal	0.45	0.25	0	1
Prob. to reach C if given as goal	0.80	0.20	0	1
Target grade	3.50	1.00	1	5
Number of study partners named	2.51	0.87	0	3
Peer incentives				
Number of study partners with incentives	1.17	0.94	0	3

Note: All control variables come from the baseline survey, except for baseline performance and teacher IDs (administrative records) and the number of study partners who were offered performance incentives (randomized after the baseline survey).

Table A.2: Balancing tests

	(1) Regression 1	(2)	(3) Regression 2	(4)
Outcome variable	Any incentive	Exogenous high	Exogenous low	Endogenous
Baseline performance				
Micro: course points (std.)	0.000	−0.044	−0.075	0.055
Controls from the survey				
Monthly spending (1,000 SEK)	−0.279	−0.211	0.118	−0.495*
Micro: attendance	−2.016	−5.444	0.582	−1.679
Micro: study with others (%)	−0.385	2.944	−1.655	−1.295
Micro: study hours	0.560	0.643	−0.037	0.799
Micro: found interesting	0.063	0.041	−0.000	0.102
Micro: satisfied with grade	−0.008	−0.119	0.106	−0.011
Prospective major: Economics	−0.008	−0.044	−0.041	0.024
Prospective major: Business Administration	0.002	−0.049	0.084	−0.013
Prospective major: Accounting	−0.006	−0.001	−0.012	−0.005
Prospective major: Philosophy & Economics	−0.016	−0.035**	−0.022	−0.004
Prospective major: Political Sciences	0.008	0.057	−0.041	0.008
Prospective major: Other	0.019	0.071	0.031	−0.009
Prob.: continue at School of Econ & Mgmt	0.810	−1.597	−2.690	3.522
Willingness to take risk	−0.182	−0.374	0.010	−0.184
Patience	−0.023	−0.101	0.151	−0.069
Self control: resist temptation	0.024	−0.059	0.308**	−0.071
Self control: am lazy	0.087	0.067	−0.017	0.145
Self control: do bad things if fun	−0.007	−0.049	−0.098	0.054
Self control: iron self-discipline	−0.035	−0.164	0.181	−0.077
Self control: long-term orientation	0.013	0.031	0.194	−0.079
Self control: want to study more	0.048	0.292*	−0.101	0.008
Optimism: expect the best	0.192*	0.050	0.290	0.210
Pessimism: can go wrong	−0.154	−0.068	0.033	−0.281**
Optimism: future	0.113	0.084	0.268*	0.054
Success: importance	0.030	0.024	0.284**	−0.086
Afraid of failure	−0.105	−0.259	−0.122	−0.029
High grade will help	−0.006	0.023	0.021	−0.033
Expected grade	0.166*	0.196	0.166	0.153
Expected grade if assigned A-goal	0.008	−0.074	−0.037	0.067
Expected grade if assigned C-goal	0.090	0.039	0.086	0.114
Prob. to reach A if given as goal	0.001	0.033	−0.006	−0.010
Prob. to reach C if given as goal	0.014	0.014	0.033	0.004
Target grade	0.133	0.147	0.168	0.110
Number of study partners named	−0.001	0.124	−0.108	−0.008
Peer incentives				
Number of study partners with incentives	−0.035	−0.019	−0.118	−0.004

Note: The table shows the results of balancing tests based on OLS regressions. Each row corresponds to two regressions. In regression 1, we regress the outcome variable on an indicator receiving any incentive. In regression 2, we regress the outcome on three treatment indicators (exogenous-high, exogenous-low, and endogenous treatment), controlling for the stratifying variables (gender, baseline performance, chosen goal), cohort and teacher IDs. The control group is the reference category in both regressions. The total number of outcome variables in this table is 36, resulting in 72 regressions and 144 coefficients displayed.

*p<0.1, **p<0.05, ***p<0.01. The p-values are computed based on cluster-robust standard errors.

Table A.3: Descriptive statistics of survey outcomes (disaggregated)

Variable	(1) Mean	(2) SD	(3) Min	(4) Max	(5) Obs.
Target grades					
Target grade (midline 1)	3.62	0.99	1	5	389
Target grade (midline 2)	3.55	1.05	1	5	381
Target grade (midline 3)	3.43	1.13	1	5	380
Effort					
Study hours (midline 1)	21.83	11.76	0	60	389
Study hours (midline 2)	22.36	11.45	0	60	381
Study hours (midline 1)	28.50	13.30	0	60	380
Study hours (endline)	40.14	13.97	0	60	383
Attendance (%, endline)	77.34	25.26	0	100	383
Study with others (%, endline)	37.28	27.92	0	100	383
Followed study plan (endline)	2.86	0.95	1	5	383
Focus (endline)	2.94	1.02	1	5	383
Motivation					
Interest (midline 1)	2.91	0.62	1	4	389
Interest (midline 2)	2.86	0.67	1	4	381
Interest (midline 3)	2.98	0.68	1	4	380
Interest (endline)	3.02	0.69	1	4	383
Continue at School (mid 1)	41.59	31.17	0	100	389
Continue at School (mid 2)	40.62	31.70	0	100	381
Continue at School (mid 3)	40.28	31.66	0	100	380
Continue at School (end)	40.05	32.18	0	100	383
Fun (endline)	3.09	1.13	1	5	383

Note: The table presents summary statistics of all survey items used to construct the survey outcomes in Section 4.3. The survey items come either from one of the three midline surveys or from the endline survey. To construct the indices, we first standardize all variables (mean zero, standard deviation one), then average across all items that belong to a topic (target grades, effort, or motivation), and then standardize again (mean zero, standard deviation one). We also test the robustness of our results to using a principle component analysis to construct the indices. See Appendix 5 for details.

Table A.4: Treatment effects on target grades

Dep.var.	(1)	(2)	(3)	(4)	(5)	(6)
	Target grade A			Target grade $\geq$ C		
	midline 1	midline 2	midline 3	midline 1	midline 2	midline 3
(T1) Exogenous-high	0.54***	0.55***	0.51***	0.06	0.04	0.14***
	(0.06)	(0.07)	(0.07)	(0.04)	(0.05)	(0.05)
(T2) Exogenous-low	0.04	–0.00	–0.03	0.08**	0.11***	0.14***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)
(T3) Endogenous	0.05	0.06*	0.03	0.08**	0.11***	0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
R-squared	0.60	0.57	0.55	0.31	0.35	0.34
Control mean	0.12	0.11	0.11	0.88	0.82	0.76
Observations	389	381	380	389	381	380
Pre-registered controls	✓	✓	✓	✓	✓	✓

Note: The table shows OLS regression results of regressions of target grades reported in the three midline surveys on treatment effect indicators (see equation 2). The outcomes are dummies for a target grade of A and a target grade of at least C. All regressions contain all pre-registered control variables. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.5: Course grades and performance in subsequent courses

Dep.var.	(1) Course grades (binary)	(2)	(3) Performance in subsequent courses (std.)	(4)
	A	At least C	Intern. Econ.	Finance
Panel A. Effect of any incentive				
Any treatment (pooled)	–0.06**	–0.07	–0.02	–0.07
	(0.02)	(0.05)	(0.08)	(0.10)
R-squared	0.25	0.27	0.45	0.41
Panel B. Effects of the different incentives				
(T1) Exogenous-high	–0.03	–0.13*	–0.15	–0.20
	(0.04)	(0.07)	(0.14)	(0.17)
(T2) Exogenous-low	–0.08***	–0.00	–0.10	–0.12
	(0.03)	(0.07)	(0.12)	(0.17)
(T3) Endogenous	–0.06**	–0.08	0.07	0.01
	(0.03)	(0.06)	(0.10)	(0.11)
R-squared	0.25	0.28	0.46	0.42
Control mean	0.08	0.48	0.00	0.01
Observations	379	379	353	289
Pre-registered controls	✓	✓	✓	✓

Note: The table shows OLS regression results on reaching an A or at least a C in the Macroeconomics course as well performance in the courses International Economics and Finance on treatment indicators. All regressions contain all pre-registered control variables. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.6: Main results: Robustness

	(1)	(2)	(3)	(4)
	Dep.var.: Course performance (std.)			
Panel A. Effect of any incentive				
Any incentive	−0.14	−0.16*	−0.20**	−0.19**
	(0.11)	(0.09)	(0.09)	(0.09)
R-squared	0.00	0.35	0.47	0.46
Panel B. Effects of the different incentives				
(T1) Exogenous-high	−0.29*	−0.27*	−0.30**	−0.32**
	(0.17)	(0.14)	(0.14)	(0.14)
(T2) Exogenous-low	−0.08	−0.14	−0.17	−0.14
	(0.15)	(0.12)	(0.12)	(0.13)
(T3) Endogenous	−0.08	−0.11	−0.17*	−0.16
	(0.12)	(0.10)	(0.10)	(0.10)
R-squared	0.01	0.36	0.47	0.46
Control mean	0.08	0.08	0.09	0.08
Observations	379	379	379	379
No controls	✓			
Reduced set of controls		✓		
Pre-registered controls			✓	✓
Without peer controls				✓

Note: The table shows results of OLS regressions of performance in the Macroeconomics course on a treatment indicator for receiving any incentive (Panel A, see equation 1) and on indicators for the different treatment arms (Panel B, see equation 2). The basic controls include baseline grades, strata, semester, and teacher indicators. The peer controls include the number of study partners and the number of study partners with incentives. Table A.1 displays the full set of preregistered controls. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.7: Robustness: Selection into exam-taking

	(1) Took exam (binary)	(2) Course performance (std.)
Panel A. Effect of any incentive		
Any treatment (pooled)	–0.00	–0.17*
	(0.03)	(0.09)
R-squared	0.23	0.39
Panel B. Effects of the different incentives		
(T1) Exogenous-high	0.06*	–0.33**
	(0.03)	(0.15)
(T2) Exogenous-low	0.04	–0.08
	(0.04)	(0.12)
(T3) Endogenous	–0.05	–0.15
	(0.04)	(0.10)
R-squared	0.24	0.40
Control mean	0.90	0.08
Observations	425	379
Pre-registered controls	✓	✓
IPW (exam-taking)		✓

Note: The table presents results of OLS regressions of participation in the exam and course performance on treatment indicators. In both regressions we control for all pre-registered control variables. In column 2 we use inverse probability weighting (IPW) in addition (Wooldridge, 2007). We weigh the observations as follows: First, we regress a dummy variable for exam participation on treatment indicators (3 treatment indicators), baseline performance (linear), and an interaction between the treatment indicators and baseline performance (linear and squared term). We extract the fitted values from this regression for each observation (individual). We then weight each observation by the inverse of the predicted probability. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.8: Robustness: Additional sessions

	(1)	(2)
	Dep.var.: Course performance (std.)	
Sample:	Full sample	Without extra sessions
Panel A. Effect of any incentive		
Any treatment (pooled)	–0.20**	–0.22**
	(0.09)	(0.09)
R-squared	0.47	0.47
Panel B. Effects of the different incentives		
(T1) Exogenous-high	–0.30**	–0.30**
	(0.14)	(0.15)
(T2) Exogenous-low	–0.17	–0.18
	(0.12)	(0.13)
(T3) Endogenous	–0.17*	–0.19*
	(0.10)	(0.10)
R-squared	0.47	0.47
Control mean	0.08	0.08
Observations	379	361
Pre-registered controls	✓	✓

Note: The table presents results of OLS regressions of course performance on treatment indicators. In column 2, we exclude students who did not fill in the baseline survey during class. In both regressions, we control for all pre-registered control variables. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01. *p<0.1, **p<0.05, ***p<0.01.

Table A.9: No effects of peers' treatment

	(1)	(2)
	Dep.var.: Course performance (std.)	
(T1) Exogenous-high	−0.33**	−0.31**
	(0.15)	(0.15)
(T2) Exogenous-low	−0.13	−0.14
	(0.12)	(0.12)
(T3) Endogenous	−0.15	−0.15
	(0.10)	(0.10)
At least one peer: exogenous-high		0.18
		(0.11)
At least one peer: exogenous-low		0.05
		(0.09)
At least one peer: endogenous		−0.02
		(0.09)
R-squared	0.42	0.42
Control mean	0.08	0.08
Observations	379	379
Pre-registered controls	✓	✓

Note: The table shows OLS regression results. The dependent variable is course performance (macro points). Column one shows the main results. In column 2, we control for whether students named at least one peer in the baseline survey who was later assigned to any of the treatment conditions. Control mean: mean of the dependent variable in the control group. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.10: Heterogeneity: Time spent studying with others at baseline

	(1)	(2)
	Dep.var.: Course performance (std.)	
Heterogeneity:	By fraction of time spent studying w/ others	
	low	high
Panel A. Effect of any incentive		
Any incentive	−0.19	−0.18
	(0.15)	(0.11)
R-squared	0.44	0.46
Panel B. Effects of the different incentives		
(T1) Exogenous-high	−0.52*	−0.20
	(0.27)	(0.18)
(T2) Exogenous-low	−0.05	−0.15
	(0.21)	(0.15)
(T3) Endogenous	−0.11	−0.18
	(0.16)	(0.13)
Control mean	0.06	0.10
R-squared	0.45	0.46
Observations	172	207
Pre-registered controls	✓	✓

Note: The table presents results of OLS regressions of course performance on treatment indicators. The sample is grouped by the fraction of study time spent studying with others as reported at baseline, using a median split (below median: 0–39%; above median: 40–100%). In both regressions, we control for all pre-registered control variables. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.11: Goal Mismatch by Target Grade

	(1)	(2)	(3)
	Dep.var.: Course performance (std.)		
	Mismatch by target grade		
Mismatched	−0.36***	−0.48**	−0.25
	(0.13)	(0.22)	(0.21)
Matched	−0.06	−0.28	0.05
	(0.14)	(0.21)	(0.23)
Sample	all	target grade	
		$\leq$ C	$\geq$ B
Observations	269	132	137
R-squared	0.52	0.57	0.64
Control mean	0.08	−0.16	0.32
Pre-registered controls	✓	✓	✓

Note: The table shows OLS regression results of the treatment effect of matched and mismatched treatments. We define a student as "matched" if the student reported a target grade of at least B at baseline and received the exogenous-high treatment, or if they reported a target grade of at most C at baseline and received the exogenous-low treatment, and as "mismatched" if the student reported a target grade of at least B at baseline and received the exogenous-low treatment, or if they reported a target grade of at most C at baseline and received the exogenous-high treatment. In all regressions, we exclude students in the endogenous treatment group and we control for all pre-registered controls. Clustered standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.12: Survey results: Motivation and effort

Dep.var.:	(1) Motivation	(2) Effort
Panel A. Effect of any incentive		
Any treatment (pooled)	0.06	0.15*
	(0.07)	(0.09)
R-squared	0.65	0.41
Panel B. Effects of the different incentives		
(T1) Exogenous-high	–0.06	0.22
	(0.11)	(0.14)
(T2) Exogenous-low	–0.02	0.05
	(0.10)	(0.14)
(T3) Endogenous	0.17*	0.16*
	(0.08)	(0.10)
R-squared	0.66	0.41
Control mean	–0.07	0.03
Observations	361	361
Pre-registered controls	✓	✓

Note: The table shows results of OLS regressions of survey measures of motivation and effort on treatment indicators. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.13: Effects on study hours per week

	(1)	(2)	(3)
	Dep.var.: Study hours		
Survey:	Midline 1	Midline 2	Midline 3
Panel A. Effect of any incentive			
Any treatment (pooled)	0.86	2.38**	0.90
	(1.00)	(1.05)	(1.34)
R-squared	0.43	0.38	0.26
Panel B. Effects of the different incentives			
(T1) Exogenous-high	2.61*	2.43	0.82
	(1.58)	(1.62)	(2.21)
(T2) Exogenous-low	0.34	4.17**	1.28
	(1.75)	(1.80)	(1.78)
(T3) Endogenous	0.30	1.51	0.74
	(1.11)	(1.14)	(1.56)
R-squared	0.43	0.38	0.26
Control mean	21.83	21.36	28.54
Observations	348	343	344
Pre-registered controls	✓	✓	✓

Note: The table presents results of OLS regressions on study hours, elicited in each of the three midline surveys, on treatment indicators. In all regressions we control for all pre-registered control variables. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

ONLINE APPENDIX (NOT FOR PUBLICATION)

Appendix A: Aggregated survey measures (target grades, effort, motivation)

We used several survey items from the midline surveys and the endline survey to construct measures of target grades, effort, and motivation.

Target grades

We used the following survey items from the midline surveys:

Q2.3 Midline surveys: What is your target grade for the course Makroekonomi (Letter scale A–E, letters are recoded as numbers to get a scale from 1–5 in steps of 1, with 5 corresponding to an A). *The question was identical in all three midline surveys, and we use the responses from all surveys.*

To construct the dependent variable *target grades*, we standardized each variable (mean of zero and standard deviation one) in the first step. In the second step, we summed up all standardized variables and then standardized the sum to have a mean of zero and a standard deviation of one.

To check the robustness of our results, we extracted the first principle component from a PCA of the standardized variables instead.

Effort

We used the following survey items from the midline surveys and the endline survey:

Q2.1 Midline surveys: During the past 7 days, how many hours do you think you have studied for the course Makroekonomi (Scale 0–60 in steps of 1). *The question was identical in all three midline surveys, and we use the responses from all surveys.*

Q3.2 Endline survey: What percentage of lectures did you attend in the course Makroekonomi? (Scale 0–100 in steps of 1)

Q3.3 Endline survey: What percentage of time that you spend studying for the Makroekonomi exam did you study together with other Makroekonomi students? (Scale 0–100 in steps of 1)

Q3.4 Endline survey: On average, how many hours per week did you study Makroekonomi during the Makroekonomi course? (Scale 0–60 in steps of 1)

Q3.9 Endline survey: How focused were you while studying during the Makroekonomi course compared to Mikroekonomi? (Scale 0–5 in steps of 1)

Q3.10 Endline survey: How much did you follow your study plans during the Makroekonomi course compared to Mikroekonomi? (Scale 1–5 in steps of 1)

To construct the dependent variable *effort*, we standardized each variable (mean of zero and standard deviation one) in the first step. In the second step we summed up all standardized variables, and then standardized the sum to have a mean of zero and a standard deviation of one.

To check the robustness of our results, we extracted the first principle component from a PCA of the standardized variables instead.

Motivation

We used the following survey items from the midline surveys and the endline survey:

Q2.2 Midline survey: How interesting is it to study the content of the course Makroekonomi? (Scale 1–4 in steps of 1). *The question was identical in all three midline surveys, and we use the responses from all surveys.*

Q2.3 Midline survey: How likely is it that you are going to study Nationalekonomi after this semester? (Scale 0–100 in steps of 1). *The question was identical in all three midline surveys, and we use the responses from all surveys.*

Q3.5 Endline survey: How interesting was it to study the content of the course Makroekonomi? (Scale 1–4 in steps of 1)

Q3.7 Endline survey: How likely is it that you are going to continue studying Nationalekonomi after this semester? (Scale 0–1 in steps of 1)

Q3.8 Endline survey: How fun was it to study during the Makroekonomi course compared to Mikroekonomi? (Scale 1–5 in steps of 1)

To construct the dependent variable *effort*, we standardized each variable (mean of zero and standard deviation one) in the first step. In the second step we summed up all standardized variables, and then standardized the sum to have a mean of zero and a standard deviation of one.

To check the robustness of our results, we extracted the first principle component from a PCA of the standardized variables instead.

Appendix B: Experimental instructions

We added comments on the design in italics. These comments were not visible to the students.

B.1 Information and consent

The AIM study

We are Erik Wengström and Pol Campos-Mercade, members of the research group in Behavioral and Experimental Economics at the School of Economics and Management at Lund University. We are carrying out an investigation that studies different ways to motivate university students in their studies. This project uses economic incentives to reward students who attain academic goals and studies their effects on the students' academic performance. The research project is generously funded by The Crafoord Foundation.

We hereby invite you to participate in the **AIM (Academic Interest and Motivation) Study**. If you decide to participate, we will ask you to fill out one survey today (about 15 minutes), three small surveys on the 8th, 15th, and 22nd of November (about 30 seconds each), and one last survey after your Makroekonomi exam (about 5 minutes). In total, **we expect that it will take you less than 25 minutes to fill out all the surveys**. In the surveys we will ask for your name, personal number, e-mail address, Swish/bank number, and questions related to your study habits, your study colleagues, your personality and your preferences.

By participating, you will both contribute to science and to the research carried out at Lund University. To compensate you for your time, **if you fill out all the surveys you will be paid 250 kr in December**.

In addition, about 60% of the participants will be offered to be paid between 2,000 and 4,000 kr depending on their Makroekonomi grade. You will get more information during the survey.

The information that you provide us in the survey will be linked to the grades of the courses that you have previously studied at Lund University and to the grades that you get in the course Nationalekonomi: Grundkurs. The data will strictly follow the GDPR 2016/679 EU regulations on data protection. The information will be completely confidential and only accessible by the two researchers responsible for this study, Pol Campos-Mercade and Erik Wengström. The data will not be used for any other purpose than this study. In March 2020, once the study is over, the data will be completely anonymized (meaning that there will not be any single data point that makes it able to identify you as an individual, such as your name, personal number, or your e-mail). Only when the data is completely anonymized may it be shared with other researchers, always with research purposes only.

Lund University is responsible to protect these data. You have the right to obtain a registry extract and correct any incorrect data concerning you. The participation is voluntary and you can decide to quit the study at any moment. For any issues related to data protection you can contact Jerker Holm, the Head of the Department of Economics, Lund University via the e-mail research.studies@nek.lu.se. If you decide to participate, please fill out the following consent:

Please, do not talk or interact with other students from now on until you leave the classroom. If you have any questions, raise your hand and wait for any of the researchers to come and talk to you directly.

B.2 Baseline questionnaire

Part one

Q1.1 Please fill out the following information about yourself.
Answer: blank field for Name, Surname, personnummer

Q1.2 Your gender
Answer: Male / Female / Other

Q1.3 How much do you spend on average every month (in kronor)? (Do not include room, food, nor transport. Include everything else. For example: study material, parties, holidays, cinemas, restaurants, mobile phones, etc.)
Answer: Blank field

Q1.4 What percentage of lectures did you attend in the course Mikroekonomi? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%.)
Answer: Scale of 0–100 in steps of 1.

Q1.5 What percentage of the time that you spent studying for the Mikroekonomi exam did you study together with other Mikroekonomi students? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%).
Answer: Scale of 0–100 in steps of 1.

Q1.6 On average, how many hours per week did you study Mikroekonomi during the Mikroekonomi course? (Count everything (study, homework, exercises), but do not count lectures. Select 60 if you studied more than 60 hours per week).
Answer: Scale of 0–60 in steps of 1.

Q1.7 How interesting was it to study the content of the course Mikroekonomi?
Answer: Very interesting / Interesting / Uninteresting / Very uninteresting

Q1.8 How satisfied are you with the grade that you got in the course Mikroekonomi?
Answer: Very satisfied / Satisfied / Unsatisfied / Very unsatisfied

Q1.9 Who is your main teacher in the course Makroekonomi?
Answer: Name of teacher 1 / Name of teacher 2 / Name of teacher 3

Q1.10 What is your target grade for the course Makroekonomi?
Answer: E / D / C / B / A

Q1.11 What program do you study or do you plan to study? (If you are unsure, answer the one that you believe is more likely.)

Answer: Economics / Business Administration / Accounting / Philosophy & Economics / Political Science / Other (blank field)

Q1.12 How likely is it that you are going to continue studying Nationalekonomi after Nationalekonomi:Grundkurs? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%.
Answer: Scale of 0–100 in steps of 1.

Q1.13 Now think about the three study colleagues that you have studied the most with during the Mikroekonomi course. In the questions below, select their name. (Note: If you for example had two study colleagues, select their name in the first two questions and do not select any name in the last question. Note 2: If there is some colleague that you cannot find, select your other colleagues and move forward.)

Select the name of your first study colleague. Note: The names are ordered alphabetically according to the first name.
Answer: Dropdown menu of all students enrolled in Mikroekonomi

Select the name of your second study colleague.
Answer: Dropdown menu of all students enrolled in Mikroekonomi

Select the name of your third study colleague.
Answer: Dropdown menu of all students enrolled in Mikroekonomi

Q1.14 How do you see yourself: are you a person who is generally willing to take risks, or do you try to avoid taking risks? Select a number 0–10 where 0 means that you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks."
Answer: Scale of 0–10 in steps of 1.

Q1.15 How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? Select a number 0-10 where 0 means that you are "completely unwilling to do so" and a 10 means you are "very willing to do so." Answer: Scale of 0–10 in steps of 1.

Q1.16 Using the scale provided, please indicate how much each of the following statements reflects how you typically are. Mark the number, where 1 is "Not at all" and 5 is "Very much."
Answer scale for each item: 1–5 in steps of 1.

- I am good at resisting temptation.
- I am lazy.
- I do certain things that are bad for me, if they are fun.
- People would say that I have iron self-discipline.
- I am able to work effectively toward long-term goals.
- I would like to study more than I currently do.

- In uncertain times, I usually expect the best.
- If something can go wrong for me, it will.
- I'm always optimistic about my future.
- It is very important for me to be successful.
- I am very afraid of failure.
- Getting a high grade in Nationalekonomi:Grundkurs will help me in my future career.

Part two

In this second half of the survey, we will give you information about this study. Please, read the following lines carefully. We will then ask you questions to make sure that you understood it correctly.

We will ask you to choose one of the following Goals:

- Goal A. You get 4000 kr if your Makroekonomi grade is: A.
- Goal C. You get 2000 kr if your Makroekonomi grade is: C, B or A.

If you receive a Goal, we will communicate it to you within the next 5 days, so that you can take it into account when you have to study. For example, if you receive Goal C, in 5 days you will know that if you reach at least a C in Makroekonomi you will get 2000 kr in December.

1. About 40% of the participants will not receive any Goal.
2. About 30% of the participants will receive the Goal that they choose.
3. About 30% of the participants will receive one of the two goals randomly.

If you receive a Goal, you can only reach it if you take and pass the main exam (ordinarie tentan) in November. If you reach the Goal by taking a re-take exam (omtentan), it will not count.

Example: Imagine that you receive Goal A. Then, in the next 5 days we will e-mail you and tell you that if your Makroekonomi grade is A, you will be paid 4000 kr. Your final payment will then be 4250 kr (4000 kr for reaching the goal + 250 kr for participating in this study) if your Makroekonomi grade is A. If, instead, your Makroekonomi grade is lower than A or you did not pass the main exam, you will get 250 kr.

Q1.17 If you receive a Goal, when will you know about it? What happens if you receive Goal A and your Makroekonomi grade in November is a C?
Answer: Today / Within the next 5 days / When I get paid after the exams

Q1.18 What happens if you receive Goal A and your Makroekonomi grade in November is a C?
Answer: I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating) / I am paid

2000 kr if I fill out all the surveys (plus 250 kr for participating) / I am only paid 250 kr for participating

Q1.19 What happens if you receive Goal C and your Makroekonomi grade in November is an A?
Answer: I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating) / I am paid 2000 kr if I fill out all the surveys (plus 250 kr for participating) / I am only paid 250 kr for participating

Remember that:

- Goal A. You get 4000 kr if your Makroekonomi grade is: A.
- Goal C. You get 2000 kr if your Makroekonomi grade is: C, B or A.

Q1.20 Please, select the grade that you think you are most likely to get in Makroekonomi depending on the Goal that you receive. For example: In the third row we ask you to choose what grade you think you would get if you received Goal C. If you believe that you would most likely get a B, then choose B in this row.
Answer for each item: F / E / D / C / B / A

- You do not receive any Goal
- You receive Goal C
- You receive Goal A

Q1.21 Select the probability that you will reach every one of the Goals if you receive it. For example: if you believe that if you are given Goal A you will reach it with 40% probability, then in the first row you should select 40%.
Answer for each item: Scale of 0–100% in steps of 1%.

- Probability that I will get an A if I receive Goal A
- Probability that I will get an A, B, or C if I receive Goal C

Your choice

Now you have to decide whether you choose Goal A or Goal C. It is very important that you think carefully and you choose your favorite Goal, since it is likely that you will receive the Goal that you choose now.

Remember that:

- Goal A. You get 4000 kr if your Makroekonomi grade is: A.
- Goal C. You get 2000 kr if your Makroekonomi grade is: C, B or A.

Q1.22 Choose your goal
Answer: Goal A / Goal C

E-mail

You are almost done!

Now we only need you to confirm your e-mail. After you write down your e-mail address and click to finish the survey, you will receive an e-mail. You will have to click on the e-mail's link and then select your favorite option to get paid.

Once you have done that, you will officially become a study participant!

Q1.22: Blank fields for the most used email address, confirmation of the most used email address, alternative email address

B.3 Disclosure of the goals

Two days after the baseline survey, the treatments were disclosed to the students by e-mail. There were five different e-mails, one for the control group, one for the exogenous-high treatment, one for the exogenous-low treatment, and two different e-mails for the endogenous treatment, depending on whether the student chose the A-goal or the C-goal at baseline. Screenshots of these e-mails are reproduced below.

Figure B.1: E-mail to the Control Group

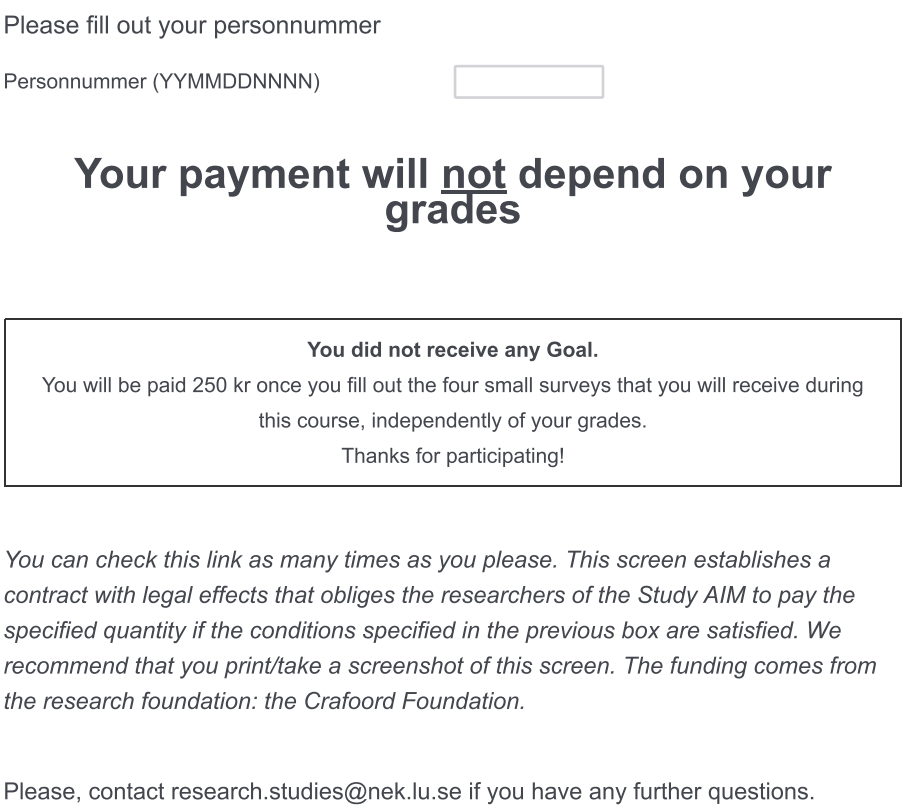


Figure B.2: E-mail to the Exogenous-High Treatment Group

Please fill out your personnummer

Personnummer (YYMMDDNNNN)

A computer randomly assigned you to one of the two goals.
You received:

GOAL A

You will be paid 4000 kr if your Makroekonomi grade is an A. This grade must be obtained by taking the Makroekonomi main exam on April 26th.

If you reach at least an A, the 4000 kr payment will be in addition to the 250 kr that you will be paid once you fill out the small surveys. This means that you will be paid 4250 kr if you reach an A.

Thanks for participating!

You can check this link as many times as you please. This screen establishes a contract with legal effects that obliges the researchers of the Study AIM to pay the specified quantity if the conditions specified in the previous box are satisfied. We recommend that you print/take a screenshot of this screen. The funding comes from the research foundation: the Crafoord Foundation.

What actions do you plan to take to make sure that you reach your goal?

You received Goal A. What happens if your Makroekonomi grade after the ordinarietentan is an A?

- I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

You received Goal A. What happens if your Makroekonomi grade after the ordinarietentan is a D?

- I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

Please, contact research.studies@nek.lu.se if you have any further questions.

Figure B.3: E-mail to the Exogenous-Low Treatment Group

Please fill out your personnummer

Personnummer (YYMMDDNNNN)

A computer randomly assigned you to one of the two goals.
You received:

GOAL C

You will be paid 2000 kr if your Makroekonomi grade is a C. This grade must be obtained by taking the Makroekonomi main exam on April 26th.

If you reach at least a C, the 2000 kr payment will be in addition to the 250 kr that you will be paid once you fill out the small surveys. This means that you will be paid 2250 kr if you reach a C.

Thanks for participating!

You can check this link as many times as you please. This screen establishes a contract with legal effects that obliges the researchers of the Study AIM to pay the specified quantity if the conditions specified in the previous box are satisfied. We recommend that you print/take a screenshot of this screen. The funding comes from the research foundation: the Crafoord Foundation.

What actions do you plan to take to make sure that you reach your goal?

You received Goal C. What happens if your Makroekonomi grade after the ordinarietentan is a B?

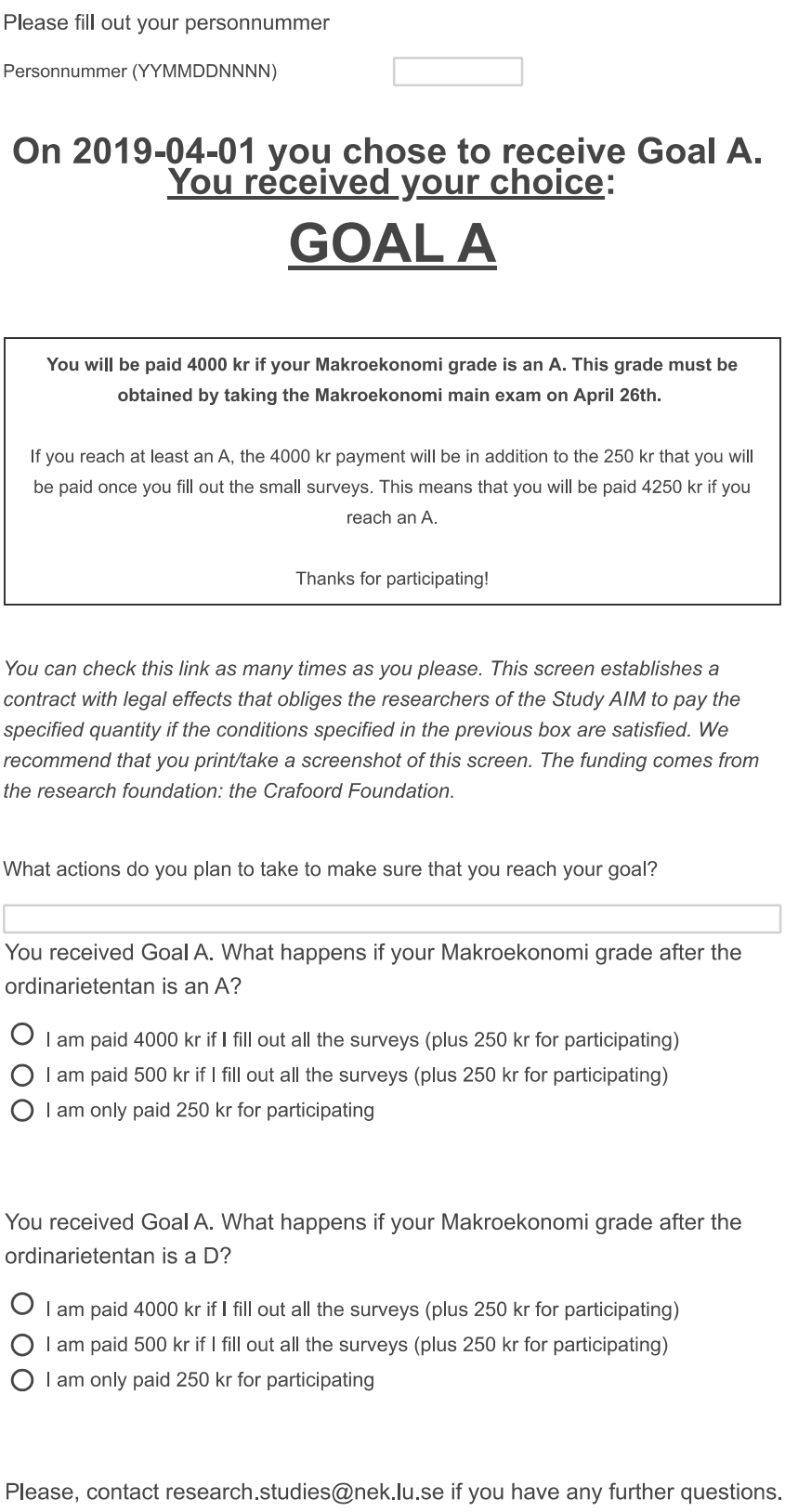
- ○ I am paid 2000 kr if I fill out all the surveys (plus 250 kr for participating)
- ○ I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- ○ I am only paid 250 kr for participating

You received Goal C. What happens if your Makroekonomi grade after the ordinarietentan is a D?

- ○ I am paid 2000 kr if I fill out all the surveys (plus 250 kr for participating)
- ○ I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- ○ I am only paid 250 kr for participating

Please, contact research.studies@nek.lu.se if you have any further questions.

Figure B.4: E-mail to the Endogenous Treatment Group with Choice of High Goal at Baseline



Please fill out your personnummer

Personnummer (YYMMDDNNNN)

On 2019-04-01 you chose to receive Goal A.
You received your choice:

GOAL A

You will be paid 4000 kr if your Makroekonomi grade is an A. This grade must be obtained by taking the Makroekonomi main exam on April 26th.

If you reach at least an A, the 4000 kr payment will be in addition to the 250 kr that you will be paid once you fill out the small surveys. This means that you will be paid 4250 kr if you reach an A.

Thanks for participating!

You can check this link as many times as you please. This screen establishes a contract with legal effects that obliges the researchers of the Study AIM to pay the specified quantity if the conditions specified in the previous box are satisfied. We recommend that you print/take a screenshot of this screen. The funding comes from the research foundation: the Crafoord Foundation.

What actions do you plan to take to make sure that you reach your goal?

You received Goal A. What happens if your Makroekonomi grade after the ordinarietentan is an A?

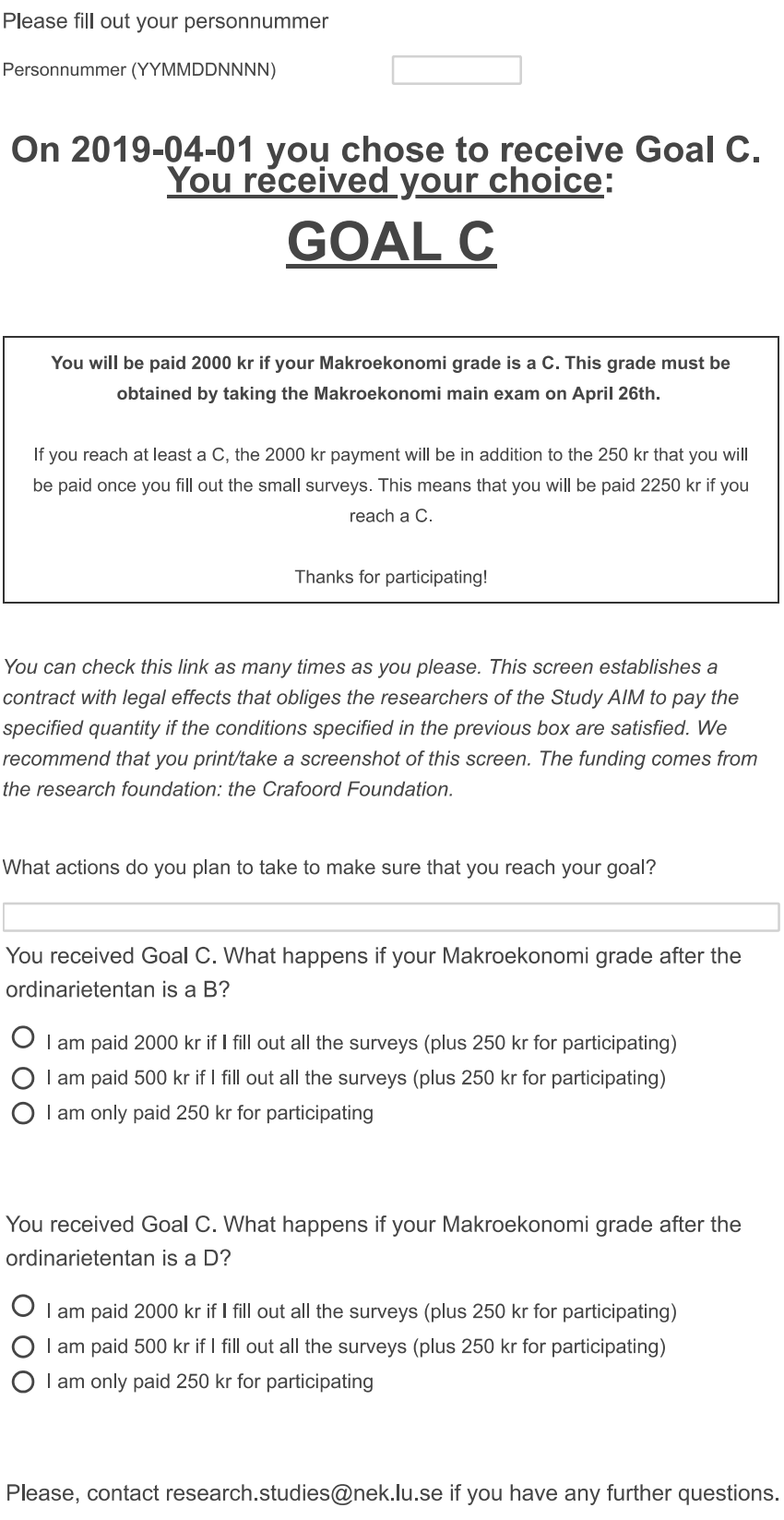
- I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

You received Goal A. What happens if your Makroekonomi grade after the ordinarietentan is a D?

- I am paid 4000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

Please, contact research.studies@nek.lu.se if you have any further questions.

Figure B.5: E-mail to the Endogenous Treatment Group with Choice of Low Goal at Baseline



Please fill out your personnummer

Personnummer (YYMMDDNNNN)

On 2019-04-01 you chose to receive Goal C.
You received your choice:

GOAL C

You will be paid 2000 kr if your Makroekonomi grade is a C. This grade must be obtained by taking the Makroekonomi main exam on April 26th.

If you reach at least a C, the 2000 kr payment will be in addition to the 250 kr that you will be paid once you fill out the small surveys. This means that you will be paid 2250 kr if you reach a C.

Thanks for participating!

You can check this link as many times as you please. This screen establishes a contract with legal effects that obliges the researchers of the Study AIM to pay the specified quantity if the conditions specified in the previous box are satisfied. We recommend that you print/take a screenshot of this screen. The funding comes from the research foundation: the Crafoord Foundation.

What actions do you plan to take to make sure that you reach your goal?

You received Goal C. What happens if your Makroekonomi grade after the ordinarietentan is a B?

- I am paid 2000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

You received Goal C. What happens if your Makroekonomi grade after the ordinarietentan is a D?

- I am paid 2000 kr if I fill out all the surveys (plus 250 kr for participating)
- I am paid 500 kr if I fill out all the surveys (plus 250 kr for participating)
- I am only paid 250 kr for participating

Please, contact research.studies@nek.lu.se if you have any further questions.

B.4 Midline questionnaire

The midline questionnaires were identical in all three midline surveys.

Q2.1 During the past 7 days, how many hours do you think you have studied for the course Makroekonomi? (Count everything (study, homework, exercises), but do not count lectures. Select 60 if you studied more than 60 hours per week).
Answer: Scale of 0–60 in steps of 1.

Q2.2 How interesting is it to study the content of the course Makroekonomi?
Answer: Very interesting / Interesting / Uninteresting / Very uninteresting

Q2.3 What is your target grade for the course Makroekonomi?
Answer: E / D / C / B / A

Q2.4 How likely is it that you are going to continue studying Nationalekonomi after this semester? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%.
Answer: Scale of 0–100 in steps of 1.

B.5 Endline questionnaire

This is the last survey to finish your participation in the AIM study. We expect it to take about 2 minutes to fill out.

Once you have filled it out, you will receive a payment of 250 kr (plus an additional payment to some of you) within the next days.

Q3.1 Write your personal number
Answer: blank field

Q3.2 What percentage of lectures did you attend in the course Makroekonomi? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%.)
Answer: Scale of 0–100 in steps of 1.

Q3.3 What percentage of the time that you spent studying for the Makroekonomi exam did you study together with other Mikroekonomi students? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%).
Answer: Scale of 0–100 in steps of 1.

Q3.4 On average, how many hours per week did you study Mikroekonomi during the Mikroekonomi course? (Count everything (study, homework, exercises), but do not count lectures. Select 60 if you studied more than 60 hours per week).
Answer: Scale of 0–60 in steps of 1.

Q3.5 How interesting was it to study the content of the course Makroekonomi?
Answer: Very interesting / Interesting / Uninteresting / Very uninteresting

Q3.6 What is your target grade for the course Nationalekonomi:Grundkurs?
Answer: E / D / C / B / A

Q3.7 How likely is it that you are going to continue studying Nationalekonomi after this semester? (Select a number from 0 to 100, where 0 is 0% and 100 is 100%.
Answer: Scale of 0–100 in steps of 1.

Q3.8 How fun was it to study during the Makroekonomi course compared to Mikroekonomi?
Answer: Much less / Less / The same / More / Much more

Q3.9 How focused were you while studying during the Makroekonomi course compared to Mikroekonomi?
Answer: Much less / Less / The same / More / Much more

Q3.10 How much did you follow your study plans (for example, tomorrow afternoon I will study 3 hours) during the Makroekonomi course compared to Mikroekonomi?
Answer: Much less / Less / The same / More / Much more

Q3.11 How stressed were you during the Makroekonomi course compared to Mikroekonomi?
Answer: Much less / Less / The same / More / Much more

Q3.12 Ten hypothetical scenarios appear below. For each, please indicate whether you would "accept" the lottery for a chance of winning (or losing) or "reject" it and not receive anything.
Answer for each item: Accept / Reject

- If the coin turns out heads, then you lose 20 kr, if it turns out tails, you win 60 kr
- If the coin turns out heads, then you lose 30 kr, if it turns out tails, you win 60 kr
- If the coin turns out heads, then you lose 40 kr, if it turns out tails, you win 60 kr
- If the coin turns out heads, then you lose 50 kr, if it turns out tails, you win 60 kr
- If the coin turns out heads, then you lose 60 kr, if it turns out tails, you win 60 kr
- If the coin turns out heads, then you lose 70 kr, if it turns out tails, you win 60 kr

Q3.13 Now think about the three study colleagues that you have studied the most with during the Makroekonomi course. In the questions below, select their name. (Note: If you for example had two study colleagues, select their name in the first two questions and do not select any name in the last question. Note 2: If there is some colleague that you cannot find, select your other colleagues and move forward.)

Select the name of your first study colleague. Note: The names are ordered alphabetically according to the first name.
Answer: Dropdown menu of all students enrolled in Makroekonomi

Select the name of your second study colleague.
Answer: Dropdown menu of all students enrolled in Makroekonomi

Select the name of your third study colleague.
Answer: Dropdown menu of all students enrolled in Makroekonomi

Q3.14 If participating in the AIM Study made you spend more time studying, where did you get that time from? For example, you can say “I partied less”, “I watched less Netflix”, “I worked out less”, or simply “I don’t know.”
Answer: open field