Performance Incentives in Education:
The Role of Goal Mismatch

Pol Campos-Mercade, Petra Thiemann, Erik Wengström

June 19, 2023

Abstract

Students often face incentives to reach performance goals, for instance to receive a scholarship, enter a college, or be hired for a job. This paper uses a field experiment to study how incentives to reach performance goals affect students, whether the effects vary for students at different parts of the performance distribution, and whether allowing students to choose their own goal improves their performance. We find that incentives backfire: students offered incentives perform worse than their control counterparts. These negative effects are mainly driven by mismatched goals: the negative treatment effects are concentrated among low-ability students who are assigned a high goal and among students with high aspirations who are assigned a low goal. The effects are also negative but not statistically significant from zero when students are allowed to choose their own goal. Our results show that incentives for performance goals can harm students' performance, especially among those students for whom the goals are mismatched.

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

JEL Codes: C93, D90, I22, I23.

1 Introduction

Incentives to reach performance goals are pervasive in education settings. For example, students are expected to reach specific performance standards to receive scholarships (Scott-Clayton, 2011), enter colleges or study fields (Kirkeboen et al., 2016), and be hired for particular jobs (Hansen et al., 2021). These incentives may change students' performance through a myriad of ways, including exerting higher effort to reach the performance goal (e.g., as in Becker 1967), affecting their intrinsic motivation (Deci and Ryan 1985), or changing their reference point (Kőszegi and Rabin 2006). A comprehensive understanding of the effects of incentives on student performance is essential for policymakers to design effective and equitable policies that promote academic success and avoid unintended consequences.

One major challenge when evaluating the impact of incentives on educational performance is that students may choose courses or programs based on the incentives they face (e.g., Bandiera et al. 2013, Dohmen and Falk 2011, Lazear 2000). For instance, students who receive scholarships may opt for easier courses, rendering their performance incomparable to those who do not receive the aid (Casey et al. 2018, Cornwell et al. 2005). Furthermore, the effects of these threshold incentives can be diverse and potentially positive or negative for different students. For example, a performance goal of a C may be favorable for an E student but harmful for an A student. Without detailed data on baseline performance in the absence of incentives, policymakers may overlook the effects of incentives on specific parts of the distribution.

To overcome these limitations, our study employs a randomized control trial (RCT) design, incentivizing students to reach specific performance-based goals in a seven-week full-time course at college.[1] This eliminates the possibility of students strategically altering their course choices. Furthermore, all students completed another seven-week full-time course before our experiment, allowing us to carefully classify students according to their previous grades and analyze whether incentives affect different students differently. In addition, we test a novel treatment in which we allow students to choose their treatment, which enables us to study the policy-relevant question of whether effectiveness improves when we allow students to self-select a performance goal. Finally, we collect a rich set of additional background data that helps us to identify the potential mechanisms.

Our sample consists of 425 students enrolled in the "Introduction to Economics" course sequence at a Swedish university (Lund University), 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

---

[1]The experiment and the main analysis were pre-registered at the AEA RCT Registry on October 29, 2018 (AEARCTR-0003507).

SEK 2000 ($220 at the time) if their Macroeconomics grade is C or higher. The "exogenous-high" treatment group offers students SEK 4000 ($440) if their Macroeconomics grade is an A. Finally, the "endogenous" treatment group allows students to choose whether they want the low goal (Goal C: earning SEK 2000 if they reach a grade of C or higher) or the high goal (Goal A: earning SEK 4000 if they reach a grade of A).[2]

We find that offering incentives to reach performance goals harms students' performance. On average, students in either incentive treatment perform 0.19 standard deviations worse than students in the control group. This effect is mainly driven by students who are assigned to the exogenous-high treatment group, who perform 0.33 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 remarkable heterogeneities. By classifying students into ability groups using their grades in the previous course, we find that the adverse effects of the exogenous-high treatment group are mainly concentrated among low-ability students. Furthermore, the negative effects of the exogenous-low treatment group are concentrated among the high-ability students. Hence, our results point to the potential misalignment between performance goals and ability as the source of the adverse effects of the incentives. We further find that the negative effects of the endogenous treatment group are concentrated among high-ability students. This effect is in line with the rest of the results: high-ability students in the endogenous treatment group overwhelmingly chose the low goal (Goal C), which in turn had adverse effects for these students.

Our detailed survey allow 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 received their preferred goal performed slightly and non-significantly worse than students in the control group. However, students who did not receive their preferred goal performed 0.39 standard deviations worse than students in the control group. These results hold similarly if we only consider students in the exogenous-low and exogenous-high treatments.

The findings suggest that the backfiring effect of incentives can occur when students are offered a performance goal that is not aligned with their personal preferences and abilities. While allowing students to choose their own performance goal can improve the effectiveness of the incentives, students are often unable to select the goal that is most beneficial to them.

Our paper contributes to a large literature on the role of financial incentives on student performance at different levels of schooling. This literature can be broadly split in two categories:

---

[2]To provide two benchmarks for the size of the incentives, a full student government grant is approximately SEK 9,000–10,000 per month, meaning that the payment of the high goal corresponds to about 40% of students' monthly allowance. Student jobs pay about SEK 100–110 per hour, meaning that the paymet of the high goal corresponds to 40 hours of work in a student job.

studies that investigate payments that increase in grades or effort (Angrist et al., 2009, Angrist and Lavy, 2009, Fryer, 2011, Bettinger, 2012, Angrist et al., 2014, Blimpo, 2014, Behrman et al., 2015, Levitt et al., 2016, List et al., 2018), and studies that incentivize meeting specific target grades or "thresholds" (Leuven et al., 2010, De Paola et al., 2012, Barrow et al., 2014, Levitt et al., 2017). The present paper belongs to the latter category.

Prior literature on threshold incentives has highlighted heterogenous effects of incentivizing a specific grade target. In particular, 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.

We introduce an experimental design that enables us to systematically study the heterogenous effects of two different grade targets and associated incentive schemes on different parts of the performance distribution. Specifically, we elicit students' preferred grade target in an incentivized way prior to assigning the goal. This is to our knowledge the first study that tests such self-chosen financial incentives in the context of a high-stakes education setting.

Our findings have important implications for the design of incentive systems in education settings, suggesting that financial incentives overall may be counterproductive, particularly for low-achievers. More research on the study habits and (non-pecuniary) motivations of low-achievers seems therefore key for understanding how their performance can be improved.

2 Background and experimental design

2.1 Institutional background

Our sample consists of 425 students enrolled in the Introduction to Economics course sequence at Lund University (LU) during the academic year 2018—2019.[3] The course runs twice a year, and we performed the experiment twice: with 193 students in the Fall semester and 232 students in the Spring semester. Introduction to Economics is a full-time course sequence that is taught over an entire semester. Since most lecture halls are not large enough to accommodate all students, at the beginning of the course, students are assigned to attend the lectures of one of three teachers who teach the course in parallel.

The course consists of four courses that are taught sequentially. First, students study a full-time course in Microeconomics for seven weeks. After their final exam on the Microeconomics course, they take a full-time course in Macroeconomics for seven more weeks. Finally, after

---

[3]Lund University is one of the largest universities in Sweden, and its undergraduate programs attract students from all over the country. Introduction to Economics is a compulsory course sequence for students in the Economics and Business Administration programs and is also taken by students from other programs in the social sciences.

they take the final exam on the Macroeconomics course, they complete two smaller courses (Finance and International Economics) simultaneously for six weeks. Students get a grade ranging from F (worst) to A (best) for each of the four courses.[4]

We conducted the experiment during the Macroeconomics course. This enables us to use the Microeconomics grade to measure baseline performance (both as a control variable to increase the statistical power of the analysis and to study heterogeneous treatment effects). Furthermore, we use the grades in the Finance and International Economics courses to explore longer-term effects. Macroeconomics was taught full-time, and its final grade was determined by 85% exam performance and 15% performance in two take-home assignments.

2.2 Experimental design

Our experimental design, summarized in Figure 1, consists of one control group and three treatment groups: 40% of the students (N = 173) did not receive any financial incentive for performance (control group); 15% of the students (N = 60) were offered SEK 2000 ($220 at the time) if their Macroeconomics grade was C or higher (exogenous-low treatment); 15% of the students (N = 62) were offered SEK 4000 ($440) if their Macroeconomics grade was an A (exogenous-high treatment). Finally, 30% of the students (N = 130) could choose whether they wanted Goal C, earning SEK 2000 if they reached a grade of C or higher, or Goal A, earning SEK 4000 if they got a grade of A (endogenous treatment group). In the endogenous treatment group, 110 students chose Goal C, and 20 chose Goal A.

The timeline of the experiment was as follows. One week after students started their Macroeconomics course, one of the researchers and a research assistant entered each of the Macroeconomics classes 20 minutes before the end of the lecture. The experiment was not pre-announced, meaning that the students were not aware of the study until the end of this lecture. The students were given a consent form, which explained that the study included five surveys: one at that very moment (15 minutes), three in each of the three following weeks (30 seconds each), and one after the Macroeconomics exam (5 minutes). By filling out all surveys, they would be paid SEK 250 ($28). In addition, some students could be offered additional money based on their grades. To participate, students were asked to sign the form and enter a link on their laptop or phone browser to fill out the first survey. To the best of our knowledge, all students who attended the lecture decided to participate. They followed our guidelines to fill out the survey individually and in silence.[5]

In the first survey, which was filled out in class, students were first asked about their socio-demographic characteristics and social security number (*personnummer*). We then asked them

---

[4]This grade is determined by the points that students got in the course. These points are mainly determined in the final exam but also may include assignments. 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.

[5]Three students had to leave earlier. We allowed these students to fill out the survey at home.

questions about their study habits (e.g., attendance and interest in both the Microeconomics and the Macroeconomics courses), as well as the names of their three best study colleagues.

At the end of the baseline survey, we asked the students to indicate their preferred goal. The students could choose to receive SEK 4,000 for reaching course grade A or SEK 2,000 for getting at least course grade C. The students were instructed to select the goal carefully, as 30% of all students would receive the chosen goal, and the remaining 70% would be randomized to another goal or to the control group. The elicitation of the preferred goal was, therefore, incentive compatible; in other words, students could not benefit from hiding their true preferences. The students could neither opt out of choosing a goal nor propose an alternative goal.

Finally, we asked the participants for their e-mail addresses. Once students had completed the survey, they were asked to leave the signed consent form on the table and leave in silence.

Two days later, all participants received an e-mail indicating their assigned treatment. Students assigned to the control group were told that their payment would be independent of their grades. The other students were told whether they received Goal C or Goal A. In addition, the students were told whether they got a given goal randomly (the exogenous treatments) or because they chose it (the endogenous treatment).

In the three midline surveys, which were very short and had to be filled out weekly, students were asked to report their interest in the Macroeconomics course, their target grade, and the likelihood that they will continue to study economics after the semester. Out of the 425 participants, 389, 381, and 380 students filled out the first, second, and third midline surveys. While participation in the control group was very high (88%), participation in the treatment groups was somewhat higher (94%). There were no differences in participation across the different treatment groups.

The endline survey, which was sent five days after the students took the final exam (but before they knew their grades), contained the same questions. In addition, it also included questions on how much fun and stress students had during the course, as well as to what extent they followed their study plans and were focused while studying. Participation was 86% in the control group and 93% in the treatment groups, with no notable differences between the control groups.

Once students received their final grades, we used each student's social security number to match them to the administration registries at the university. We then carried out the payments based on their treatment and actual grade. Finally, after students finished the entire Introduction to Economics course sequence, we matched our data with the final Microeconomics, Macroeconomics, Finance, and International Economics grades for all 425 students in the study.[6] The consent form and surveys that students received are displayed in Appendix B.

---

[6] While students can retake each of these courses several months later, we were very clear that we would only consider their grades in the main exam.

3 Data and descriptive statistics

Our data come from the five different survey waves (baseline, midline 1–3, and endline) and from administrative student records provided by the university administration. Our main sample for the analysis consists of all students who participated in the experiment. For these students, we have non-missing values for all questions in the baseline survey since participation in the experiment was conditional on completing the baseline survey. We provide additional analyses for the subsamples of students who filled in the midline and endline surveys.

3.1 Pre-assignment data

3.1.1 Baseline achievement and student characteristics

We collected performance at baseline based on the performance in the Microeconomics course. In addition, we elicited students' characteristics in a pre-assignment (baseline) survey. These characteristics included socio-economic variables (gender, spending), behavioral characteristics (willingness to take risk, patience), prospective major choices (economics, business administration, accounting, philosophy and economics, political sciences, other). We also elicited the preferred goal and asked the students about their aspirations for the course (grade target), and their probability of reaching either of the goals under each of the treatments.

Table 1 presents summary statistics of the sample. The students are majority male (56%), and most of them intended to major either in Business Administration (39%), Economics (27%), or Political Sciences (20%). Students were generally active in their studies: 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. Regarding prior performance, students received, on average, a grade between D and C in the Microeconomics course. Most students reached at least a C (57% of students), and 6% reached an A. The students were generally eager to improve upon their grades at baseline: 15% of the students 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.

3.1.2 Preferred goal

Toward the end of the survey, the students indicated their preferred goal; 14% of the students (N = 60) chose the high goal (Goal A: reward of SEK 4,000 for reaching grade A), and 86% of the students (N = 365) chose the low goal (Goal C: reward of 2,000 for reaching grade C or better).

The data indicate that the students chose their goals carefully, that is, in line with their ambition, ability, and willingness to take risk elicited at baseline (see Panels A–C of Figure 2). Nearly half of the students who received grade A at baseline chose the high goal (48%), whereas

only 21% of students who received grade B at baseline chose the high goal, and less than 10% of students who received an F–C at baseline chose the high goal. Similarly, 60% of students who had a target grade of A chose the high goal, whereas only 12% of students who had a target grade of B chose the high goal, and less than 5% of students who had a target grade of E–C chose the high goal. Reassuringly, out of those students who indicated that their target grade was C, 99% chose the low goal, which would incentivize a grade of C. Finally, a higher willingness to take risk is strongly associated with the choice of the high goal (see Figure 2, Panel C).

3.1.3 Beliefs

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

We find that the students believed that the incentive would increase their performance. On average, the expected grades were lowest when given no goal (slightly lower than C), higher when given a C-goal (between C and B), and highest when given the A-goal (close to B). No student expected the C-goal to backfire (i.e., to lead to a lower grade than no goal), and only seven students expected the A-goal to backfire. The students are rather confident about reaching assigned goals: on average, they believed that they would reach an A with a probability of 45% if assigned the A-goal, and that they would reach a C with a probability of 80% if assigned the C-goal. These numbers imply an average expected return of the A-goal of SEK 1,784 ($\approx$ USD 162) and an average expected return of the C-goal of SEK 1,593 ($\approx$ USD 145). The difference between the expected returns of the A-goal and the expected returns of the C-goal is positively associated with choosing the A-goal (see Panel D of Figure 2).

3.2 Balancing checks and treatment definition

Treatment assignment was random conditional on gender, points in the micro exam, and the chosen goal. The results of balancing checks are displayed in Table 2. We do not find any significant differences in any background characteristics across students in any treatment and the control group, neither when pooling all treatments nor when analyzing the different treatment arms (exogenous-high, exogenous-low, endogenous). This suggests that the randomization was successful in balancing student background characteristics.

3.3 Outcome variables

Our primary outcome variable are the points in the macro exam. In the exam, the students had to answer five open-ended questions about the lecture content (no multiple-choice questions). Out of the 425 students who participated in the survey, 379 students took the exam (89%). Students could reach between 0–100 points in the exam; in practice, the points ranged between 8 and 97 with a mean of 60 points. We standardize the macro points (mean zero and unit standard deviation). In addition to the points in the Macroeconomics exam, we also study the exam grades. In total, 5% of the students reached an A in the Macroeconomics exam, and 39% of the students reached a C in the Macroeconomics exam (see Panel A of Table 3).

For our supplementary analyses, we use self-reported behaviors and perceptions from the endline survey, elicited after the exam had been taken but before the grades were disclosed (see Panel B of Table 3). These survey measures are available for 90% of the sample. Finally, we also use longer-term achievement measures based on the grades in the courses Finance (available for 83% of the sample) and International Economics (available for 83% of the sample) (see Panel C of Table 3).

4 Results

4.1 Effect of incentives on Macroeconomics grades

4.1.1 Main results

We find that incentives, overall, have a negative effect on macro points and grades. Figure 3 displays the distribution of macro points (top panel) and macro grades (bottom panel) for all students in the control group and students who received either of the three treatment groups (exogenous-high, exogenous-low, or endogenous). On average, students in the control group perform 0.13 standard deviations better on the exam (corresponding to 2.1 points out of the 100 possible points) but this raw difference is imprecisely measured (p-value of Wilcoxon rank sum test = 0.20; see Table 4).

The losses accrue across the entire grade distribution. Students who received any incentive are less likely to receive a top grade (grades A–B), more likely to receive a low grade or to fail the exam (grades F–D). These effects are unlikely to be driven by selection into taking the exam, as students who receive any incentive are only one percentage point less likely to attend the exam, compared to the control group (p-value of Wilcoxon rank sum test < 0.82; see Table 4).

To formally examine the relationship between the Macroeconomic grade and the incentives, we conduct an OLS regression of the form:

$$y_{ic} = \alpha_0 + \alpha_1 incentive_i + X_i'\gamma + \delta_c + \epsilon_i, \tag{1}$$

where y_{ic} denotes the grade of student i in cohort c, *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 pre-registered control variables (gender, baseline grade, classroom and teacher dummies, peer incentives, study behavior and target grades at baseline), δ_c is a cohort dummy (Spring or Fall term), and ϵ_i is an error term.

Table 5 presents the results. Conditional on the full set of controls, incentives decrease macro points by -0.19 standard deviations on average (SD = 0.09, see column 7). Without control variables, we find somewhat lower and less precisely measured estimates. In sum, our results suggest that that incentives do not improve performance but backfire instead.

[Table 5 about here.]

Next, we investigate the effects of the three different treatment arms separately, by performing an OLS regression that includes dummy variables for the three different treatments, *exo_high$_i$* (exogenous-high treatment), *exo_low$_i$* (exogenous-low treatment), and *endo$_i$* (endogenous treatment), with the students in the control group as the omitted category:

$$y_{ic} = \beta_0 + \beta_1 exo\_high_i + \beta_2 exo\_low_i + \beta_3 endo_i + X_i'\gamma + \delta_c + \epsilon_i. \tag{2}$$

The negative effect of the financial incentives is mainly driven by the exogenous-high reward scheme (see Table 5, Panel B). In the specification with the full set of controls, students who receive the exogenous-high treatment score on average to 0.33 lower on the Macroeconomics exam, compared to the control group (SD = 0.15). We do not find any significantly negative effects of the other two treatments (exogenous-low and endogenous) in any of the regression specifications, but the point estimates are negative throughout.

4.1.2   The role of goal mismatch

The negative effect of incentives appears surprising at first sight: the students perform less well in the presence of incentives, compared to the control group. The effect, however, could be driven by mismatch between the students' own goals, and the goals provided to them. For instance, a too low goal can set a low reference point, thus decreasing the performance of the most ambitious students; similarly, a too high goal can be discouraging to those who had set a less ambitious goal, for instance by crowding-out the intrinsic motivation through an extrinsic (but elusive) reward. These behavioral explanations are likely to play out differently for different types of students: a too high goal should be negative for those who aimed for a lower goal; a low goal should be negative for those who aimed for a higher goal. In general, mismatched incentives should be less effective than matched incentives.

To test for the effect of goal mismatch, we first investigate whether the different goals affect students who chose the high versus the low goal differently. Table 6 displays the treatment effects seperately for students who chose a low and for those who chose a high goal at baseline. We also formally test for effect heterogeneity by interacting the treatments with dummies for students' chosen goal.

[Table 6 about here.]

Our results are suggestive of mismatch effects: students who chose the the low goal at baseline have a significantly negative and sizeable treatment effect of receiving the exogenous-high treatment. They perform on average -0.42 test score standard deviations worse, compared to the control group (SE = 0.16, column 2 of Table 6). By contrast, students who chose a high goal at baseline do not experience any negative effect of the exogenous-high treatment. Their treatment effect of receiving the exogenous-high treatment amounts to 0.40 test score standard deviations, but is imprecisely measured (SE = 0.75, column 3 of Table 6). In the interacted model the treatment effect of the exogenous-high treatment is significantly more positive for students who chose the high goal (difference = 0.95, SE = 0.41, column 4 of Table 6). By contrast, we do not find any effect heterogeneity for the exogenous-low and the endogenous goals.

We also provide a direct test of the effect of goal mismatch. We define a student as "matched" if the chosen goal corresponds to the randomly assigned goal, and as "mismatched" if the assigned goal is different from the chosen goal. Whether a student is matched or mismatched is random given our experimental design.[7]

We perform an OLS regression of the Macroeconomics grade on indicator variables for receiving a matched treatment, $matched_i$, and for receiving a mismatched treatment $mismatched_i$, with students in the control group being in the omitted category:

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

We find that a mismatched treatment decreases student performance by -0.39 standard deviations on average, and the effect is statistically significant (SE = 0.15, see Table 7 column 1). A matched treatment also negatively affects the students, compared to the control group, but the effect is smaller and statistically indistinguishable from zero (effect of -0.13 test score standard deviations, SE = 0.09, see column 1). The mismatch effect is robust to excluding the students who received the endogenous treatment and are thus always matched (Table 7, column

---

[7]To see that, first, consider a student who chose the high goal. 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 the low goal 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 probility of 40%.

2). The effect of mismatch appears to be less pronounced for students who chose the high goal at baseline, but we do not detect any statistically significant evidence for effect heterogeneity across students with different goals at baseline (Table 7, columns 3–5).

[Table 7 about here.]

In sum, our analysis of goal mismatch suggests that students who were assigned to their preferred goal perform better than students who are not assigned to their preferred goal; however, neither matched nor mismatched treatments raise performance relative to the control group.

4.2 Heterogeneity

4.2.1 Heterogeneity by ability

Ability differences across students could drive the effect heterogeneity we find. One might expect, for instance, that students in bottom part of the ability distribution might be discouraged by the high goal but encouraged by the low goal. We therefore explore whether the incentives help or hinder students with different abilities at baseline. We define a student as a high-ability student if she received and A and B in the Microeconomics course (i.e., at baseline), and as a low-ability student if she received a C or worse in the Microeconomics course. The rationale for this definition is that students who received a B or better might be at the margin of being motivated by the high goal.

We find that the negative treatment effect of the exogenous-high treatment is mainly driven by the low-ability students (see Table 8). The exogenous-high treatment decreases low-ability students' test scores by -0.38 standard deviations relative to the control group (SE = 0.19), but does not affect the outcomes of the high-ability students (coefficient of -0.03, SE = 0.26). We do not find, however, that the point estimates are statistically different across the two ability levels. We do not find ability differences in the effects of the the endogenous and exogenous-low treatments.

4.2.2 Heterogeneity by willingness to take risks

The endogeous treatment does not deliver any positive effects in our analysis, and the point estimate for the effect is always negative. This raises the question of whether – and why – the students chose a goal (i.e., incentive scheme) that does not suit them. A potential channel might be risk aversion. Reaching the C-goal is less risky, even for the low-ability students. Thus, risk-averse students might choose a goal that is too easy for them and subsequently adjust their effort downwards to match the chosen goal.

In Table 9 we display the results separately for students with high and low willingness to take risk. Willingness to take risk is measured on a scale of 0–10, and we define low willingness

to value of five or lower. We find a significantly negative effect of the endogenous treatment in the sample of risk-averse students. This suggests that risk-averse students are be more prone to choosing a goal that is too low and subseqently decrease their performance. However, we cannot confirm that any of the treatment effects are significantly different across the samples of students with high and low willlingness to take risks.

4.2.3 Gender heterogeneity

Finally, we investigate gender heterogeneity because men often respond more strongly to financial incentives than women. We find that our treatment effects are more pronounced for the male students, which is in line with these results from the prior literature (Table 10). We cannot confirm, however, that the treatment effects are significantly different between male and female students.

4.3 Intermediate and longer-term outcomes

4.3.1 Target grades

A potential concern in our study might be that the students misunderstood the incentives or did not care about them. To mitigate this concern, we asked the students to report their target grade in each of the three midline surveys. We can therefore observe how the students' grade targets developed over time as a function of the different incentives that the students received.

We find that the students' target grades respond to the incentives in intuitive ways, suggesting that the students understood and account for the incentives (Figure 4). We also find – in line with our main results – that the target grades respond more strongly to the exogenous-high goal than to the endogenous end exogenous-low goals. In the exogenous-high treatment, the probability of reporting a target grade of A increases by 50 percentage points, suggesting that nearly half of the students in the exogenous-high treatment group adjust their target grades upward towards an A (Panel A). This effect materializes in the first week after receiving the incentives and persists throughout all the midline surveys. By contrast, we do not find any significant effect of the exogenous-low and the endogenous treatment on the target grades, except in the last midline survey (Panels B and C). Even in this survey, the magnitude of the effect is small (less than three percentage points). This is because 89% of the students already had at least a target grade of C at baseline (see Table 1). Therefore, the scope to adjust the target grades upward to a C is limited. We also do not find strong evidence that the students adjust their target grade downwards when assigned the exogenous-low and endogenous treatment.

4.3.2 Study behaviors

Based on data from the endline survey, we investigate whether self-reported study behaviors change in response to the treatments. The investigated study behaviors are hours spent studying, lecture attendance, studying with others, and following a study plan. We do not find that any of these behaviors are affected by the treatments. In particular, we see close to null responses in the exogenous-high treatment group, relative to the control group (see Figure 5).

4.3.3 Longer-run outcomes

Finally, we investigate the effect on three longer-term outcomes: the self-reported probability of specializing in economics after completing the first year, and the grades from subsequent courses from administrative course records. None of the longer-run effects are economically large or statistically significant; that is, the negative effects did not meaningfully persist beyond the incentivized course (see Figure 6).

5   Discussion

This study investigates the effect of two different financial incentive schemes on student performance in higher education. The first incentive schemes targets high achievers in that it incentivizes the top grade and is associated with a high reward. The second incentive scheme targets students in the middle of the performance distribution in that it incentivizes a mid-level grade and is associated with a smaller reward. To improve the match between the incentive scheme and students' ability and aspirations, we also assign to some students their preferred incentive scheme.

The incentives backfire, and the negative effect stems mostly from those students who received the higher goal. Assigning the self-chosen goals is also ineffective at raising performance. Low-ability students are particularly harmed by receiving a goal that does not match their aspirations.

Our findings caution against the use of financial incentives when the student body is diverse. While we cannot exclude that some students benefit from the incentives, the overall effect is either zero or detrimental. Moreover, offering a menu of goals does not automatically lead to better results. Instead, more research is needed to understand the inner workings of students' motivation and study habits in order to design treatments that target improvements in behaviors and study skills.
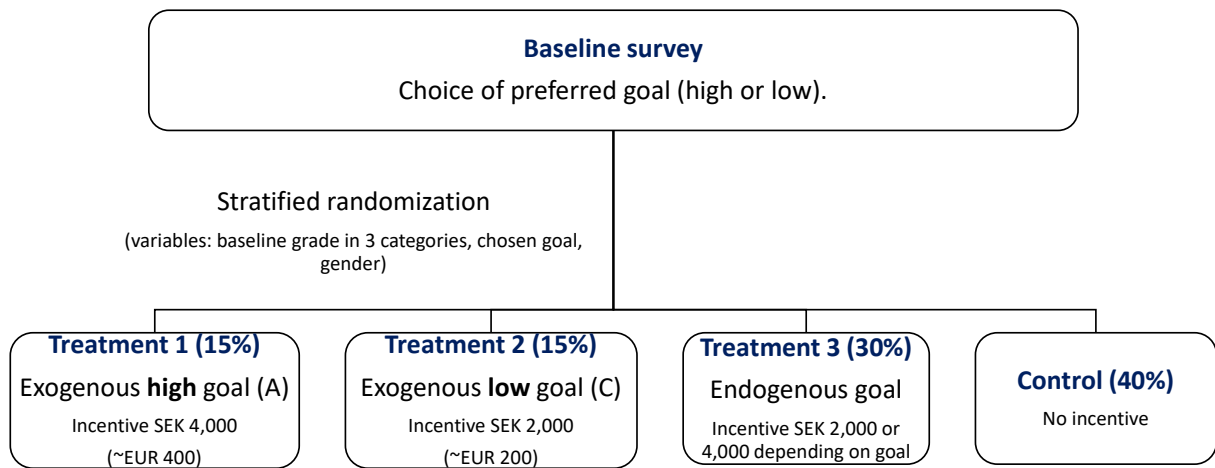
References

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. (2014). When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources*, 49(3):572–610.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5):1079–1114.

Barrow, L., Richburg-Hayes, L., Rouse, C. E., and Brock, T. (2014). Paying for Performance: The Education Impacts of a Community College Scholarship Program for Low-Income Adults. *Journal of Labor Economics*, 32(3):536–599.

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.

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.

Blimpo, M. P. (2014). Team Incentives for Education in Developing Countries A Field Randomized Experiment in Benin. *American Economic Journal: Applied Economics*, 6(4):90–109.

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.

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.

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.

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

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

Hansen, A. T., Hvidman, U., and Sievertsen, H. H. (2021). Grades and employer learning.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.

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

Lazear, E. P. (2000). 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*.

List, J. A., Livingston, J. A., and Neckermann, S. (2018). Do financial incentives crowd out intrinsic motivation to perform on standardized tests? *Economics of Education Review*, 66:125–136.

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.

Figure 1: Experimental design

```
┌─────────────────────────────────────────────────────────────────────┐
│                           Baseline survey                             │
│                 Choice of preferred goal (high or low).               │
└─────────────────────────────────────────────────────────────────────┘
                                    │
                        Stratified randomization
             (variables: baseline grade in 3 categories, chosen goal,
                                  gender)

┌──────────────────┐  ┌──────────────────┐  ┌──────────────────┐  ┌──────────────────┐
│ Treatment 1 (15%)│  │ Treatment 2 (15%)│  │ Treatment 3 (30%)│  │                  │
│ Exogenous high   │  │ Exogenous low    │  │ Endogenous goal  │  │  Control (40%)   │
│ goal (A)         │  │ goal (C)         │  │                  │  │  No incentive    │
│ Incentive SEK    │  │ Incentive SEK    │  │ Incentive SEK    │  │                  │
│ 4,000            │  │ 2,000            │  │ 2,000 or         │  │                  │
│ (~EUR 400)       │  │ (~EUR 200)       │  │ 4,000 depending  │  │                  │
│                  │  │                  │  │ on goal          │  │                  │
└──────────────────┘  └──────────────────┘  └──────────────────┘  └──────────────────┘
```

Note: Letters in brackets denote the incentivized grade (A in treatment 1 and C or better in treatment 2. Grades are measured on a letter scale from A–F (F: fail, A: best grade). High goal: A. Low goal: C or better. Percentages denote the fractions of students assigned to each of the treatment groups.

Figure 2: Fraction of students choosing the high goal by baseline characteristics
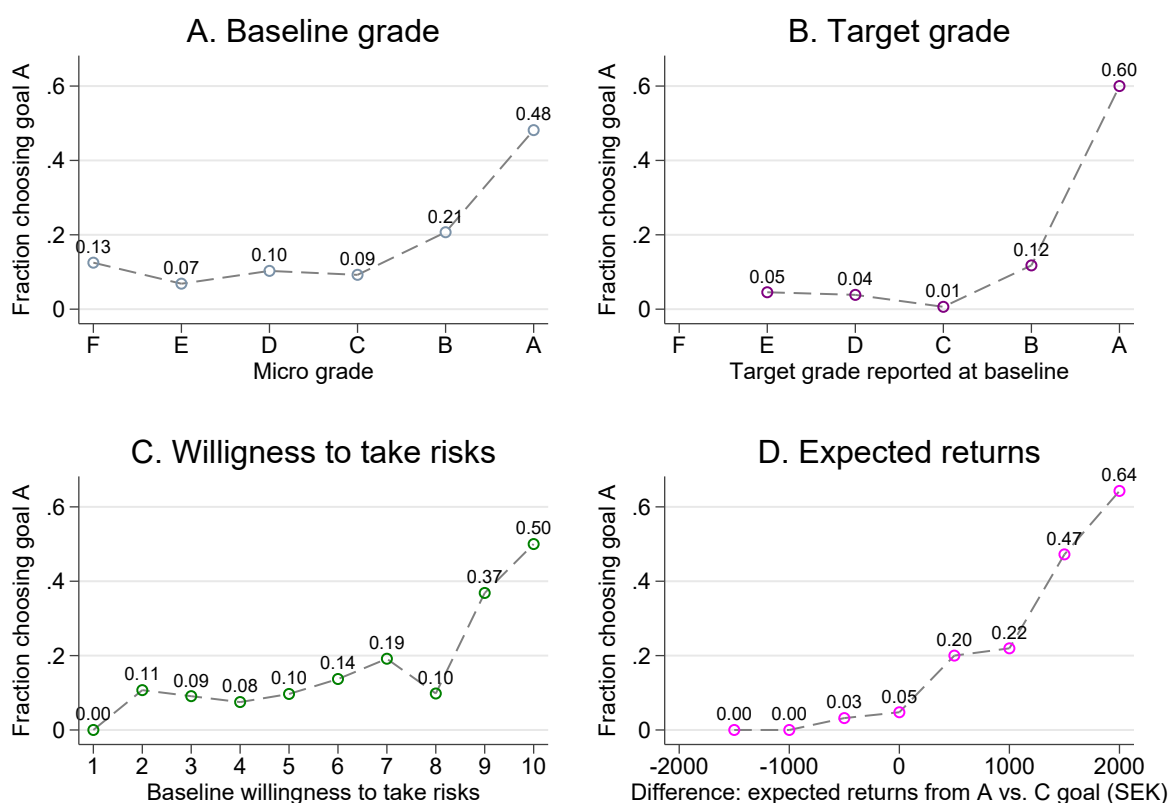
Note: The figure displays the fraction of students who chose the "high" goal (reward of SEK 4,000 for getting an A in the macro course) as their preferred reward scheme on the y-axis, and different baseline characteristics in each panel on the x-axis (baseline grade, target grade, willingness to take risk, expected monetary return from the high versus the low goal). The grades come from administrative student records, the target grade, willingness to take risk, an questions used to compute expected returns come from the pre-treatment 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 getting 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.

CDF of macro points

Effect of any incentive on grades
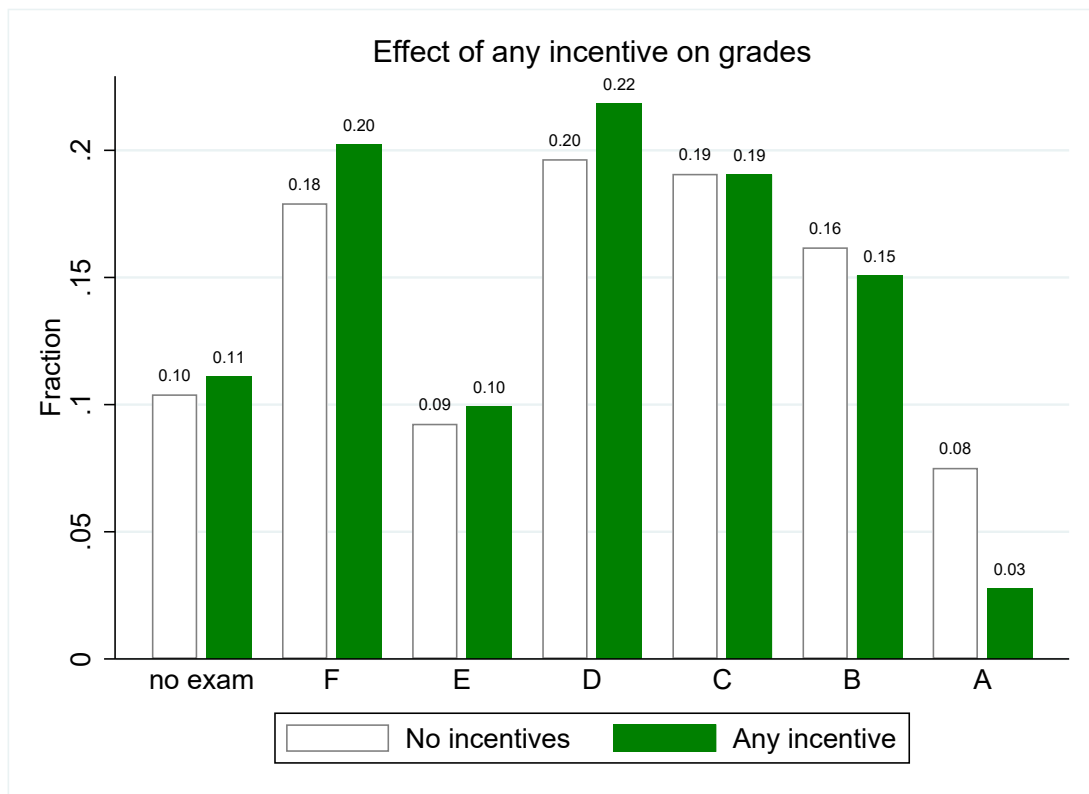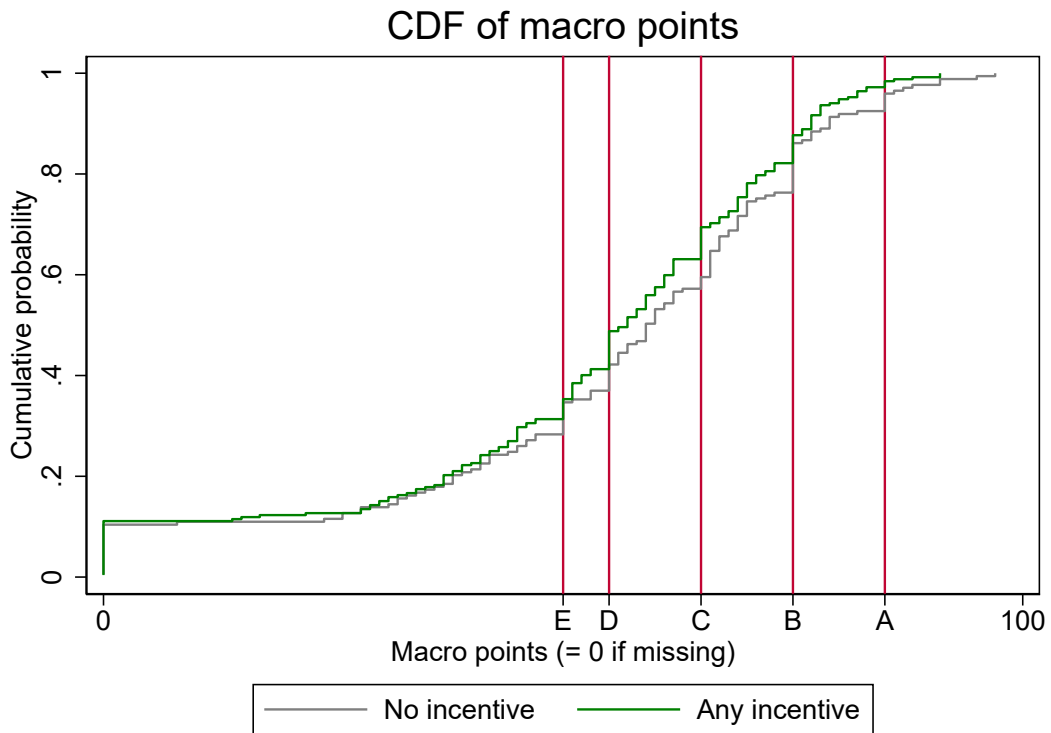
Figure 3: CDF of macro points (top panel) and histogram of macro grades (bottom panel). The graphs compare the control group (students who do not receive any incentive) to the pooled treatment group (students who receive any incentives (exogenous-high, exogenous-low, endogenous). In the top panel, students who did not attend the exam were coded as receiving zero points.

Figure 4: Effect of the treatments on target grades reported in the three midline surveys
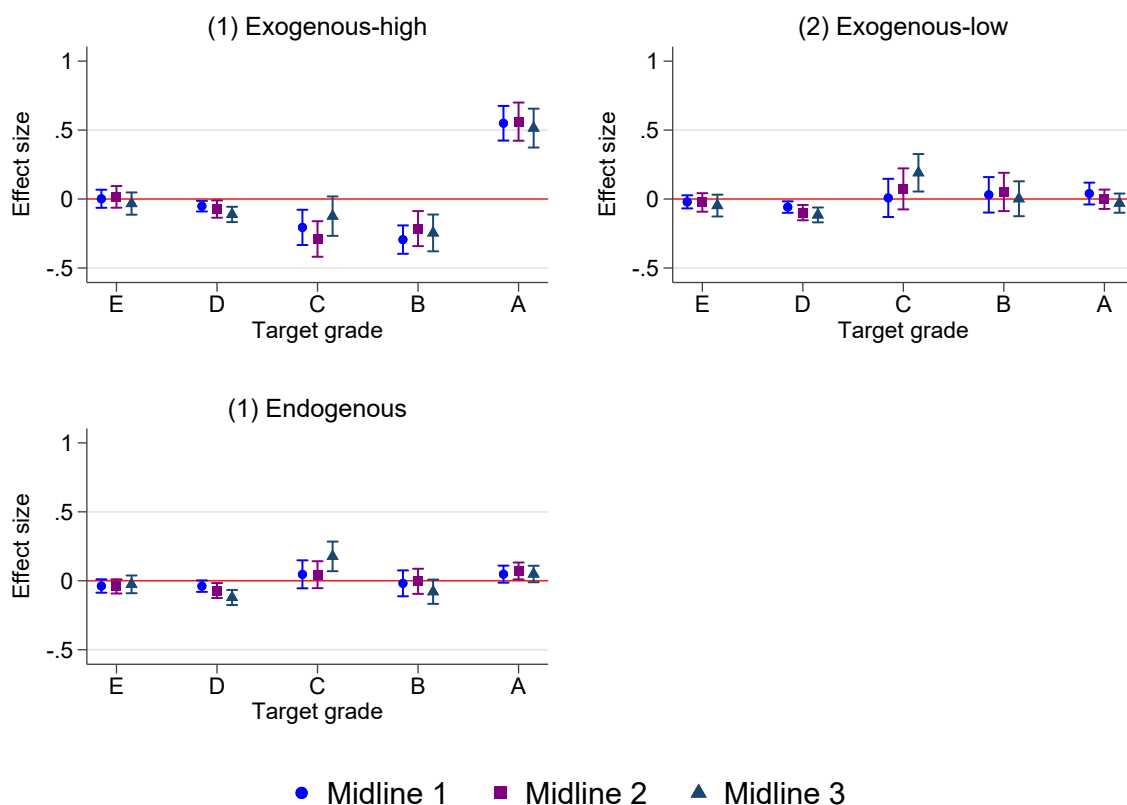
Note: The figure displays regression results of OLS regressions of target grades on treatment effect indicators. Each panel shows the results for one of the three treatments (exogenous-high, exogenous-low and endogenous). Target grades come from the three midline surveys. To compute the coefficients, we regressed the indicator variables for the respective target grade (e.g., dummy for a target grade of A) on the three treatment dummies and the full set of pre-registered control variables. This was done for each of the five target grades and each of the three surveys, resulting in fifteen regressions. The whiskers display 95%-confidence intervals.

Figure 5: Intermediate outcomes
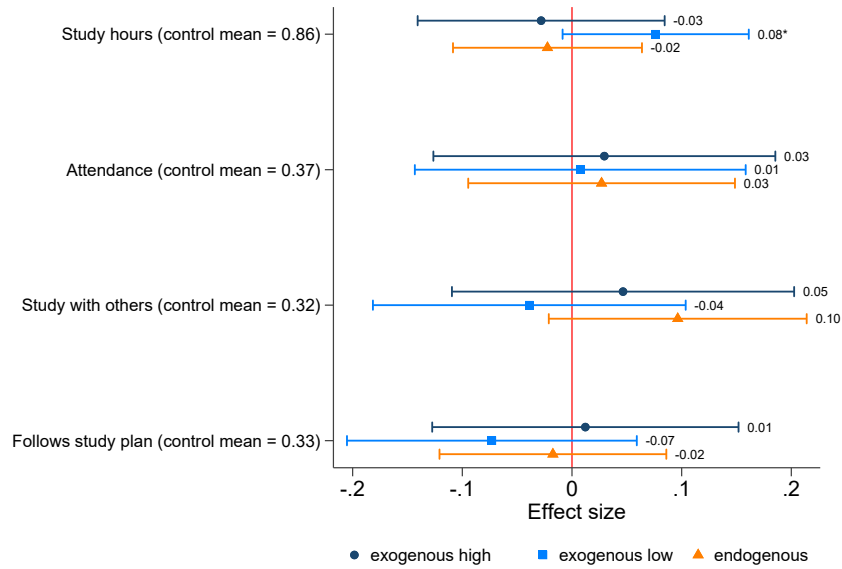
Note: The figure displays results of OLS regressions of outcomes from the endline survey on treatment indicators. The outcomes are measured as differences between the baseline and endline survey. For instance, a value of 0.86 indicates that 86% of the students increased their study hours relative to the prior course. The whiskers display 95%-confidence intervals.

Figure 6: Longer-run outcomes
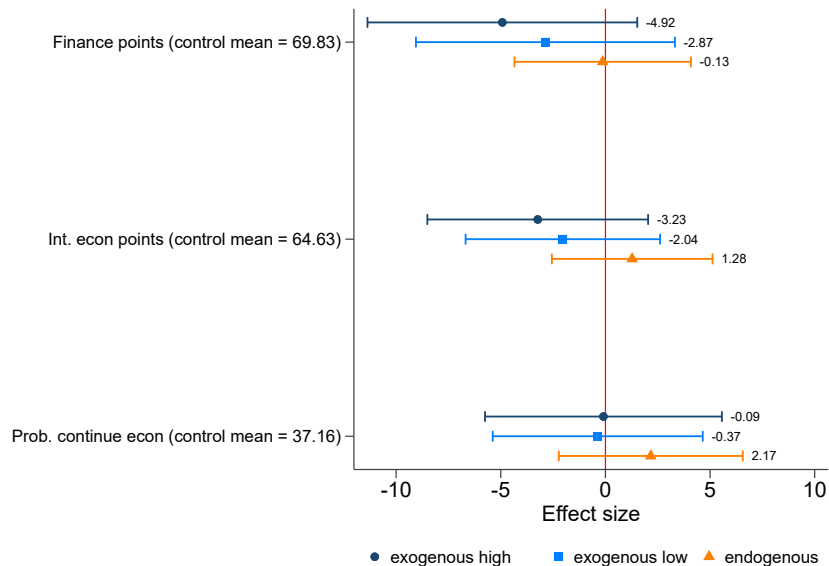
Note: The figure displays results of OLS regressions of three different longer-run outcomes (points in subsequent courses, self-reported probability to continue studying economics) on treatment indicators and the pre-registered controls. Exam points and the probability to continue in economics are reported on a scale of 1–100. The whiskers display 95%-confidence intervals.

Table 1: Baseline Survey: Summary Statistics

	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Background					
Female	425	0.44	0.50	0	1
Monthly spending[1] (1,000 SEK)	425	3.85	2.53	0	18
Study partners named	425	2.51	0.87	0	3
Willingness to take risk	425	5.69	2.04	1	10
Patience	425	7.17	1.80	1	10
Prior performance					
Microeconomics grade[2]	425	2.55	1.40	0	5
Microeconomics grade A	425	0.06	0.24	0	1
Microeconomics grade at least C	425	0.57	0.50	0	1
Micro: study hours	425	23.60	11.07	2	60
Micro: study with others (%)	425	42.83	29.49	0	100
Micro: attendance (%)	425	79.29	22.60	0	100
Micro: found interesting	425	3.03	0.64	1	4
Micro: satisfied with grade	425	2.62	0.92	1	4
Major choice					
Major: Economics	425	0.27	0.44	0	1
Major: Business adminstration	425	0.39	0.49	0	1
Major: Accounting	425	0.04	0.20	0	1
Major: Philosophy & economics	425	0.04	0.18	0	1
Major: Political sciences	425	0.20	0.40	0	1
Major: Other	425	0.07	0.25	0	1
Probability: study econ	425	44.88	30.14	0	100
Macro: aspirations					
Target grade[2]	425	3.50	1.00	1	5
Target: A	425	0.15	0.36	0	1
Target: at least C	425	0.89	0.32	0	1
Macro: beliefs					
Expected grade if given no goal[2]	425	2.87	1.02	0	5
Expected grade if given A goal[2]	425	3.92	1.03	0	5
Expected grade if given C goal[2]	425	3.49	0.69	0	5
Expectation: reach A if given as goal (%)	425	44.61	24.60	0	100
Expectation: reach C if given as goal (%)	425	79.66	19.88	0	100
Choice of goal					
Choice: goal A	425	0.14	0.35	0	1
Treatments					
(T1) Exogenous-high	425	0.15	0.35	0	1
(T2) Exogenous-low	425	0.14	0.35	0	1
(T3) Endogenous	425	0.31	0.46	0	1
Control group (no incentive)	425	0.41	0.49	0	1

Note: All variables come from the baseline survey, except for he Microeconomics grade (administrative records) and the treatments (randomized after the survey was completed).

[1]Monthly spending excludes costs of room, food, and transportation. 1 SEK corresponded to ≈ 0.11 USD at the time of the survey.

[2]Letter grades are recoded into numbers for the purpose of creating summary statistics: 0 = F, 1 = E, 2 = D, 3 = C, 4 = B, 5 = A.

Table 2: Balancing tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Female		Spending		Friends		Risk		Patience		Micro grade		Study hours		Attendance	
Any treatment (pooled)	-0.02		-291.56		-0.03		-0.16		0.00		-0.10		0.22		-2.17	
	(0.05)		(250.01)		(0.09)		(0.20)		(0.18)		(0.14)		(1.09)		(2.23)	
(T1) Exogenous-high		-0.04		-158.33		0.09		-0.29		-0.03		-0.12		0.31		-5.47
		(0.07)		(374.51)		(0.13)		(0.30)		(0.27)		(0.21)		(1.64)		(3.34)
(T2) Exogenous-low		-0.02		86.45		-0.13		-0.02		0.13		0.01		-0.27		0.64
		(0.07)		(379.08)		(0.13)		(0.31)		(0.27)		(0.21)		(1.66)		(3.38)
(T3) Endogenous		-0.02		-529.57*		-0.03		-0.16		-0.04		-0.15		0.41		-1.90
		(0.06)		(293.68)		(0.10)		(0.24)		(0.21)		(0.16)		(1.29)		(2.62)
Mean of dependent variable	.44	.44	3,854	3,854	2.5	2.5	5.7	5.7	7.2	7.2	3.5	3.5	24	24	79	79

	Interest in micro		Satisfied w/ grade		Major: econ		Major: business		Major: accounting		Major: philosophy		Major: political		Major: other	
Any treatment (pooled)	0.06		-0.05		0.00		0.03		-0.00		-0.02		-0.03		0.02	
	(0.06)		(0.09)		(0.04)		(0.05)		(0.02)		(0.02)		(0.04)		(0.02)	
(T1) Exogenous-high		0.05		-0.13		-0.01		0.01		0.01		-0.05*		-0.04		0.07*
		(0.09)		(0.14)		(0.07)		(0.07)		(0.03)		(0.03)		(0.06)		(0.04)
(T2) Exogenous-low		-0.02		0.09		-0.03		0.11		-0.01		-0.03		-0.06		0.03
		(0.10)		(0.14)		(0.07)		(0.07)		(0.03)		(0.03)		(0.06)		(0.04)
(T3) Endogenous		0.09		-0.07		0.03		-0.01		-0.00		-0.00		-0.01		-0.01
		(0.07)		(0.11)		(0.05)		(0.06)		(0.02)		(0.02)		(0.05)		(0.03)
Mean of dependent variable	3	3	2.6	2.6	.27	.27	.39	.39	.04	.04	.035	.035	.2	.2	.068	.068

	Study econ		Target grade		Target: A		Target: at least C		Grade if no goal		Expects A if goal A		Expects C if goal C		Choice: goal A	
Any treatment (pooled)	0.14		0.08		0.04		0.03		0.14		-1.11		0.32		-0.03	
	(2.98)		(0.10)		(0.04)		(0.03)		(0.10)		(2.43)		(1.96)		(0.03)	
(T1) Exogenous-high		-2.38		0.11		0.07		0.07		0.19		2.06		2.06		-0.05
		(4.46)		(0.15)		(0.05)		(0.05)		(0.15)		(3.65)		(3.65)		(0.05)
(T2) Exogenous-low		-4.33		0.07		0.02		0.03		0.09		-2.93		-2.93		-0.08
		(4.51)		(0.15)		(0.05)		(0.05)		(0.15)		(3.69)		(3.69)		(0.05)
(T3) Endogenous		3.41		0.08		0.04		0.02		0.13		-1.77		-1.77		-0.01
		(3.50)		(0.12)		(0.04)		(0.04)		(0.12)		(2.86)		(2.86)		(0.04)
Mean of dependent variable	45	45	3.5	3.5	.15	.15	.89	.89	2.9	2.9	45	45	80	45	.14	.14

Note: The table shows results of OLS regressions of baseline characteristics on treatment dummies. Based on 425 observations. Robust standard errors are in parentheses. * $p < 0.1$.

Table 3: Outcomes: summary statistics

	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
A. Performance – short-term					
Macro points	379	59.69	15.80	8	97
Macro grade[1]	379	2.18	1.53	0	5
No macro exam	425	0.11	0.31	0	1
Macro grade A	425	0.05	0.21	0	1
Macro grade C or better	425	0.39	0.49	0	1
Macro passed	425	0.70	0.46	0	1
B. Behavior & perceptions					
Macro: attendance (%)	383	77.34	25.26	0	100
Macro: study with others (%)	383	37.28	27.92	0	100
Macro: study hours	383	40.14	13.97	0	60
Macro: found interesting	383	3.02	0.69	1	4
Macro: target grade	383	3.15	1.16	1	5
Macro: more fun	383	3.09	1.13	1	5
Macro: more focus	383	2.94	1.02	1	5
Macro: better followed study plan	383	2.86	0.95	1	5
Macro: more stress	383	3.57	1.03	1	5
Probability: study econ	383	40.05	32.18	0	100
C. Performance – longer-term					
Finance points	289	69.67	18.87	4	100
Finance grade[1]	320	2.76	1.84	0	5
International economics points	353	64.59	19.13	5	99
International economics grade[1]	358	2.57	1.70	0	5

Note: All variables come from the endline survey, except for he points and grades in macro, finance, and interantional economics (administrative records). [1]Letter grades are recoded into numbers for the purpose of creating summary statistics: 0 = F, 1 = E, 2 = D, 3 = C, 4 = B, 5 = A.

Table 4: Outcomes across treatments (non-parametric tests)

	(1) Control	(2) Any incentive	(3) Exo-high	(4) Exo-low	(5) Endo-genous	(6)	(7) Differences	(8)	(9)
						(2)-(1)	(3)-(1)	(4)-(1)	(5)-(1)
Points (std.)	0.08	-0.06	-0.21	-0.00	0.00	-0.14 (0.198)	-0.29 (0.100)	-0.08 (0.600)	-0.08 (0.433)
Grade A	0.08	0.03	0.05	0.02	0.02	-0.05 (0.024)	-0.03 (0.474)	-0.06 (0.101)	-0.05 (0.045)
At least B	0.24	0.18	0.21	0.20	0.15	-0.06 (0.141)	-0.03 (0.662)	-0.04 (0.557)	-0.08 (0.075)
At least C	0.43	0.37	0.34	0.43	0.35	-0.06 (0.224)	-0.09 (0.221)	0.01 (0.940)	-0.07 (0.194)
At least D	0.62	0.59	0.53	0.63	0.59	-0.04 (0.445)	-0.09 (0.205)	0.01 (0.901)	-0.03 (0.573)
At least E	0.72	0.69	0.68	0.68	0.69	-0.03 (0.505)	-0.04 (0.560)	-0.03 (0.624)	-0.02 (0.644)
Took exam	0.90	0.89	0.95	0.92	0.85	-0.01 (0.818)	0.06 (0.188)	0.02 (0.644)	-0.05 (0.196)

Note: The table shows the means of the outcomes (points in macro exams, grades in the macro exam) in the control group (column 1), in each of the treatment arms (columns 3–5) and in all treatments pooled (column 2). Columns 7–9 display the differences between the outcomes in the treatment arms and in the control group, and column 6 between all treatments pooled and the control group. P-values of Wilcoxon ranksum tests (two-sided tests) are in parentheses.

Table 5: Treatment effects on macro points (std.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Effect of any treatment (pooled)							
Any treatment (pooled)	−0.14	−0.11	−0.14	−0.16*	−0.16*	−0.18**	−0.19**
	(0.11)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
R-squared	0.005	0.271	0.337	0.354	0.357	0.398	0.401
Panel B. Effects of the different treatments							
(T1) Exogenous-high	−0.29*	−0.25*	−0.24*	−0.27*	−0.27*	−0.32**	−0.33**
	(0.17)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)
(T2) Exogenous-low	−0.08	−0.06	−0.13	−0.14	−0.15	−0.12	−0.13
	(0.15)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
(T3) Endogenous	−0.08	−0.06	−0.09	−0.11	−0.11	−0.15	−0.16
	(0.12)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
R-squared	0.010	0.275	0.339	0.356	0.359	0.402	0.405
Observations	379	379	379	379	379	379	379
Control mean	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Strata		✓	✓	✓	✓	✓	✓
Micro points			✓	✓	✓	✓	✓
Classroom and teacher				✓	✓	✓	✓
Peer incentives					✓	✓	✓
Baseline characteristics						✓	✓
Target at baseline							✓

Note: The table shows OLS regression results. The dependent variable are macro points, standardized to have a mean of zero and a standard deviation of one. In panel A, the main regressor is an indicator for receiving any treatment. In panel B, the main regressors are treatment dummies for each type of treatment. Control mean: mean of the dependent variable in the control group. Each column contains different sets of control variables. **Strata:** dummies for each randomization stratum. **Peer incentives:** indicator for number of friends who receive an incentive. **Classroom and teacher:** teacher-by-semester dummies. **Micro points:** points in the micro course (third-order polynomial). **Baseline characteristics:** survey characteristics collected at baseline, especially those that predict attrition (monthly spending, number of study partners, risk, micro: study hours, micro: attendance, micro: found interesting, micro: satisfied with grade, dummies for intended major, probability to continue in econ). Target at baseline: target grade in macro course, target grade of A, target grade of at least C. Robust standard errors are in parenthes.
*p<0.1, **p<0.05, ***p<0.01.

Table 6: Effect heterogeneity by students' own chosen goal (Dep.var.: macro points (std.))

Sample:	(1) all students	(2) chose low goal	(3) chose high goal	(4) all students
(T1) Exogenous-high	−0.33**	−0.42***	0.40	−0.41**
	(0.15)	(0.16)	(0.75)	(0.16)
× choice: high goal				0.95**
				(0.41)
(T2) Exogenous-low	−0.13	−0.14	−0.39	−0.12
	(0.12)	(0.12)	(0.56)	(0.13)
× choice: high goal				0.00
				(0.52)
(T3) Endogenous	−0.16	−0.16	−0.47	−0.14
	(0.10)	(0.10)	(0.43)	(0.10)
× choice: high goal				−0.16
				(0.31)
Observations	379	331	48	379
R-squared	0.40	0.42	0.83	0.42
Control mean	0.08	−0.00	0.53	0.08
Full set of controls	✓	✓	✓	✓

Note: The table shows OLS regression results. The dependent variable are macro points (std.), and the main regressors are dummies for each treatment. Column 1 corresponds to Panel B., column 7 in Table 5. Columns 2 and 3 show the results by chosen goal at baseline (low vs. high). Column 3 shows the interacted model as a formal test for effect heterogeneity. Robust standard errors are in parentheses.

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

Table 7: Effect of goal mismatch (Dep.var.: macro points (std.))

Sample:	(1) all students	(2) all students	(3) chose low goal	(4) chose high goal	(5) all students
Treated, matched	−0.13	−0.07	−0.15	−0.16	−0.14
	(0.09)	(0.13)	(0.09)	(0.46)	(0.09)
× choice: high goal					0.06
					(0.29)
Treated, mismatched	−0.39**	−0.36**	−0.42***	−0.38	−0.42***
	(0.15)	(0.16)	(0.16)	(0.57)	(0.16)
× choice: high goal					0.30
					(0.51)
Observations	379	269	331	48	379
R-squared	0.408	0.423	0.422	0.800	0.409
Control mean	0.08	0.08	−0.00	0.53	0.08
Full set of controls	✓	✓	✓	✓	✓
Without endogenous treatment		✓			

Note: The table shows OLS regression results of the treatment effect of matched and mismatched treatments, where a match is defined as alignment with the chosen goal at baseline. Students in the control group are in the omitted category (neither matched nor mismatched). *p<0.1, **p<0.05, ***p<0.01.

Table 8: Heterogeneity by ability (Dep.var.: macro points (std.))

Sample:	(1) all students	(2) ability high	(3) ability low	(4) all students
(T1) Exogenous-high	−0.33**	−0.03	−0.38**	−0.35*
	(0.15)	(0.26)	(0.19)	(0.19)
× ability high				0.09
				(0.30)
(T2) Exogenous-low	−0.13	−0.20	−0.06	−0.05
	(0.12)	(0.23)	(0.15)	(0.15)
× ability high				−0.27
				(0.27)
(T3) Endogenous	−0.16	−0.18	−0.14	−0.11
	(0.10)	(0.18)	(0.12)	(0.13)
× ability high				−0.17
				(0.21)
Observations	379	112	267	379
R-squared	0.405	0.454	0.335	0.408
Control mean	0.081	0.723	−0.179	0.081
Control reach A	0.084	0.191	0.037	0.084
Control reach C	0.477	0.723	0.370	0.477
Full set of controls	✓	✓	✓	✓

Note: The table shows OLS regression results in the full sample and by baseline ability (ability high: grade of A or B in the prior course; ability low: grade lower than B or failed the prior course). The outcome are the points in the Macroeconomics exam, standardized to mean zero and standard deviation one. Robust standard errors are in parentheses.
*p<0.1, **p<0.05, ***p<0.01.

Table 9: Heterogeneity by willingness to take risks (Dep.var.: macro points (std.))

Sample:	(1) all students	(2) risk high	(3) risk low	(4) all students
(T1) Exogenous-high	−0.33**	−0.16	−0.37	−0.41**
	(0.15)	(0.21)	(0.23)	(0.21)
× risk high				0.15
				(0.26)
(T2) Exogenous-low	−0.13	−0.01	−0.31*	−0.25
	(0.12)	(0.19)	(0.18)	(0.15)
× risk high				0.21
				(0.22)
(T3) Endogenous	−0.16	−0.14	−0.29**	−0.15
	(0.10)	(0.15)	(0.14)	(0.13)
× risk high				−0.02
				(0.17)
Observations	379	211	168	379
R-squared	0.405	0.404	0.530	0.407
Control mean	0.08	0.41	0.22	0.08
Control reach A	0.08	0.07	0.11	0.08
Control reach C	0.48	0.41	0.57	0.48
Full set of controls	✓	✓	✓	✓

Note: The table shows OLS regression results in the full sample and by willingness to take risk (risk high: value of six or higher on a scale of 0–10; risk low: values lower than six). The outcome are the points in the Macroeconomics exam, standardized to mean zero and standard deviation one. Robust standard errors are in parentheses.
*p<0.1, **p<0.05, ***p<0.01.

Table 10: Heterogeneity by gender (Dep.var.: macro points (std.))

Sample:	(1) all students	(2) female	(3) male	(4) all students
(T1) Exogenous-high	−0.33**	−0.24	−0.39*	−0.52**
	(0.15)	(0.19)	(0.22)	(0.22)
× female				0.42
				(0.29)
(T2) Exogenous-low	−0.13	−0.09	−0.18	−0.16
	(0.12)	(0.19)	(0.17)	(0.16)
× female				0.07
				(0.25)
(T3) Endogenous	−0.16	−0.12	−0.17	−0.19
	(0.10)	(0.14)	(0.14)	(0.14)
× female				0.06
				(0.20)
Observations	379	171	208	379
R-squared	0.405	0.514	0.449	0.410
Control mean	0.08	0.51	0.06	0.08
Control reach A	0.08	0.07	0.10	0.08
Control reach C	0.48	0.51	0.44	0.48
Full set of controls	✓	✓	✓	✓

Note: The table shows OLS regression results in the full sample and by gender. The outcome are the points in the Macroeconomics exam, standardized to mean zero and standard deviation one. Robust standard errors are in parentheses.
*$p<0.1$, **$p<0.05$, ***$p<0.01$.

Appendix B: Experimental instructions

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

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 students were informed about the GDPR and asked to sign a consent form. Then, they accessed the baseline survey in qualtrics.*]

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.

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 10.

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 10.

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 10.

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 10

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 <u>one</u> 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 10%

- 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

3 Midline questionnaire

Q2.1 During the past 7 days, how many hours do you think you have studied for the course Macroeconomi? (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 10.

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 10

4 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 10.

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 10.

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 10.

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 10

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