The effect of financial rewards on students' achievement: Evidence from a randomized experiment*

Edwin Leuven   Hessel Oosterbeek   Bas van der Klaauw

Abstract

This paper reports about a randomized field experiment in which first year university students could earn financial rewards for passing all first year requirements within one year. Financial incentives turn out to have positive effects on achievement of high ability students, whereas they have a negative impact on achievement of low ability students. After three years these effects have increased, suggesting dynamic spillovers. The negative effects for less able students are consistent with results from psychology and behavioral economics showing that external rewards may be detrimental for intrinsic motivation.

Keywords: financial incentives, student achievement, randomized social experiment, heterogeneous treatment effects, higher education policy

JEL Codes: I21, I22, J24

1 Introduction

Recently, there is increased interest in the effectiveness of financial incentives for students to improve their achievements (e.g. Angrist et al. 2002, 2006; Angrist and Lavy 2002, 2005; Dearden et al. 2002; Kremer et al. 2004). A reason for this interest is the impression that students often do not exert the amount of study effort they are supposed to. Standard economic theory predicts a positive relation between financial incentives and achievement. Insights from behavioral economics and mixed empirical evidence do, however, cast doubt on the strength of this relation in each context and for each group of individuals.

Camerer and Hogarth (1999) review 74 studies where subjects were paid zero, small or large financial rewards for a large variety of tasks. The effects of incentives on performance in these studies are mixed and complicated. Camerer and Hogarth point to two important

*This is a revised and extended version of Leuven et al. (2003). Leuven and Oosterbeek are at the Department of Economics of the University of Amsterdam. Van der Klaauw is at the Department of Economics of the Free University in Amsterdam. The authors are affiliated with research program 'Scholar' and the Tinbergen Institute. We thank Josh Angrist, Michael Kremer, Guy Laroque, Steve Pischke, Erik Plug and seminar participants in Amsterdam, Dublin, Sevilla, Stockholm, Tilburg, and Uppsala for their comments.

factors to explain the variation in findings the importance of intrinsic motivation, and the match between what is needed to earn a reward and an individual's capabilities.

Both factors are likely to play an important role in education contexts. Students, especially those in higher education, are likely to have some degree of intrinsic motivation. For intrinsically motivated students the introduction of financial incentives may have adverse effects. They may perceive the introduction of the financial incentives as a signal the task at hand is less interesting that they thought. Different students posses different capabilities so that a given requirement to obtain the reward will be easier to fulfill for some than for others. Those for whom the requirement is clearly out of reach, will not put in more effort. The effect of financial incentives in education is therefore an empirical question.

A few other studies present experimental evidence on the effectiveness of financial incentives in the context of educational production, and on how students respond to financial rewards. Angrist and Lavy (2005) analyze the effects of financial rewards on students' achievement in an experimental setting. They evaluate the effectiveness of financial incentives that reward secondary education matriculation in Israel. They find that the intervention led to a substantial increase in matriculation rates among girls. Angrist et al. (2006) evaluate how merit-scholarships and services affect achievement at a large Canadian university. They find no effects for boys. Girls had higher grades which faded out somewhat after a year, and the treatment that combined the merit-scholarship with peer advising and study group services was more effective. Kremer et al. (2004) analyze the effects of financial rewards on achievement for primary school girls in rural Kenya by means of a randomized experiment. The experiment was conducted in two districts in western Kenya and shows large positive effects on both achievement and school attendance in one of these districts. There is also evidence for substantial externalities. Although only girls were eligible it is found that boys (who were ineligible), and girls with low initial achievement (who were unlikely to earn a reward) also experienced higher test scores and school attendance.[1]

This paper studies the effect of financial incentives on achievement and effort by means of an experiment among first year undergraduate students in economics and business at the University of Amsterdam. The experimental design assigned freshmen to three different groups. Students assigned to the large reward group could earn a bonus of NLG 1,500 (€ 681) on completion of all first year requirements by the start of the next academic year. That is, they should all 60 credit points in one year. Students assigned to the small reward group could earn a NLG 500 (€ 227) bonus for this achievement. Students who were assigned to the control group could not earn a reward. The design with both a small and a large reward allows us to separate the effect of receiving a financial reward from the

[1] Two other programs worth mentioning, although they do not have an experimental setup, are the Education Maintenance Allowance (EMA) in the United Kingdom and Columbia's PACES program. Both interventions provide financial incentives for achievement. The EMA gives low-income families a payment for enrollment and achievement. Assignment to treatment is, however, not random. Dearden et al. (2001, 2002) describe the evaluation of this program. PACES is a program in which more than 125,000 Colombian pupils received vouchers which covered about half of the cost of private secondary school. Vouchers were only renewed for pupils who maintained satisfactory academic performance (Angrist et al., 2002).

effect of the size of the reward. To examine the heterogeneity of incentive effects by student ability, the randomization was conducted in such a way that the ability distributions in the three groups are identical. To put the requirement into context: in a normal year in which no financial incentives are provided, around 20 percentage of the students passes all first year exams in one year, and the average number of collected credit points is around 30 (with a standard deviation of 22).

To briefly summarize our results, for the full sample we find a small and insignificant positive effect of the large reward on achievement, both measured by pass rates and numbers of collected credit points. This is, however, the result of two opposing effects. High ability students have higher pass rates and collect significantly more credit points when assigned to (larger) reward groups. Low ability students on the other hand appear to achieve less when assigned to the large reward group. While at the end of the first year these effects are only significant for the high ability group, after three years the sizes of the effects has increased and are statistically significant for both low and high ability students. Because the rewards were only tight to first year performance, this suggests positive dynamic spillovers. This is an important finding because it is sometimes argued that financial rewards make students work harder during the period that the incentive is in place, but they may slow down afterward. Our findings contradict such a mechanism.

The remainder of this paper is organized as follows. Section 2 provides relevant background information about the Dutch system of higher education and of the economics and business program at the University of Amsterdam. Section 3 explains the design of the field experiment and describes the data. Section 4 presents and discusses the empirical results discusses the interpretation of the findings. Section 5 discusses potential threats to the validity of the experiment such as substitution bias, manipulation by teachers, and externalities, and concludes that these are unlikely to affect our conclusions. Section 6 summarizes and discusses our findings.

2 Background

University education in the Netherlands is accessible to students with a qualification from the pre-university track in secondary education. This secondary education qualification can only be obtained by passing a uniform nationwide exam. The relevant secondary education exit requirements are set such that they are considered to be sufficient university entry requirements, and therefore all students starting a university education in economics or business are supposed to be capable of actually graduating (given that they exert sufficient effort). In the academic year 2001/2002 there were 34,200 first year students at Dutch universities, which is about 17 percent of the relevant birth cohort. Universities are not permitted to select students; everyone who applies with a valid entry qualification has to be admitted.[2] In the Netherlands selection therefore takes place at the exit of secondary

[2]For a few studies students are admitted on the basis of a lottery when the number of applicants exceeds the number of available places. This is not the case for the economics and business studies.

education as opposed to the entry of higher education.

Currently, six Dutch universities offer an undergraduate program in economics and business. While there are small differences between the programs offered by these universities, they are considered to be close substitutes. Not only do they attract students from the same pool of secondary school graduates, but they prepare their students for the same labor market, although people tend to stay in their region of origin.

University students in the Netherlands are all charged the same tuition fee (NLG 2,930 (€1,300) in the academic year 2001/2002). The tuition fee is set by the government and does not vary by field of study or by university. There is also a financial aid system which pays all university students up to NLG 1,424 (€646) per month. The financial aid scheme consists of three components (roughly equal at the maximum amount) that students are entitled to for a maximum of four years. The first component is a basic grant, the second an additional grant that decreases with parental income, and the third (optional) component is a loan.

The loan component of the financial aid scheme is not very popular among Dutch students. Students typically use the basic grant and the additional grant, but of the total amount available for loans less than 20 percent is requested. This type of debt aversion is not only observed in the Netherlands, see for example Field (2006) who finds evidence for debt aversion among law students at New York University. Instead of taking up the loan many university students combine studying with some hours of paid work. In our sample around 80 percent of the students work, and they work on average around 12 hours per week (details concerning data collection are provided later).

The undergraduate program in economics and business at the University of Amsterdam has a nominal duration of 4 years. In the first academic year, which runs from September until August, all students in economics and business follow exactly the same program of 14 compulsory courses. The first year program was divided into three terms of 14 weeks each in the year that the experiment was conducted. It is important to note that, since the program is fixed, students cannot substitute easy for difficult courses. Every term ended with exams shortly after the courses finished and the re-take exams are organized in March/April and the last week of August. The first academic year thus consisted of 42 study weeks, which are allotted to different courses in the form of 60 credit points.[3] It is only after the first term of their second academic year that students choose different packages of courses to specialize either in economics or in business.

The first year pass rate among students in economics and business at the University of Amsterdam is typically in the vicinity of 20 percent. Such low pass rates are not uncommon in continental European countries (e.g. Garibaldi et al., 2006) and can be attributed to an institutional arrangement in which universities are publicly funded and where tuition fees are low or non-existing. As a consequence, students are not confronted with appropriate prices and spend more time in the system than the nominal duration of their studies.

[3]Table A1 in the appendix gives an overview of the first year courses and the number of credit points assigned to each course.

For society, study delay imposes a cost in the form of extra expenditures on education and the foregone productivity of the students. The Department of Economics at the University of Amsterdam has an incentive to increase the pass rate since funding depends partly on the number of credits points awarded each year. There are also other reasons to address the delay of students: teaching becomes more difficult because not all students are on schedule, the failing and re-taking of exams also implies more crowded classrooms and more grading. Moreover, once a year a ranking of university departments in each field is published which is aimed at secondary education students who are in the process of choosing their university education. The first year pass rate is one of the inputs of this ranking.

At the beginning of the third trimester in the academic year 1999/2000 the low pass rate among first year students spurred the dean of the economics department to promise all undergraduate econometrics freshmen a reward of NLG 1,000 (€ 454) upon fulfilling all first year requirements before the start of the next academic year. In the Netherlands, econometrics is a separate undergraduate education from economics and business, and attracts students from the upper part of the ability distribution. In the year that this reward was in place the pass rate was 0.50 compared to 0.28 in the previous year (cf. Hilkhuysen 2000).It is difficult to establish a causal relation between the financial incentive and the increased pass rate given the non-experimental nature of the intervention. While the 0.22 increase in the pass rate may be the causal effect of the reward, this need not be the case. Plausible alternative explanations for the increased pass rate are a higher quality of the student cohort, less demanding courses, and less strict grading of exams. Nevertheless the results suggest that a financial incentive may be a very effective policy intervention.

3 Experimental design and data

To investigate the effectiveness of financial rewards to increase achievement of university students, we conducted a field experiment among first year economics and business students at the University of Amsterdam. The experiment took place in the academic year 2001/2002.

To ensure that all students were treated identically, participation in the experiment was only open to students who (i) followed the full-time program, (ii) did not claim more than 1 credit point dispensation,[4] and (iii) did not start the economics and business program in a previous year. The total number of eligible students equals 254.

Participation in the experiment was on a voluntary basis. On October 1, 2001, almost one month after classes started, we sent all first year students a letter inviting them to participate in the experiment. This was the earliest possible date given the availability of addresses from the student administration. The letter explained the purpose of the

[4]Students can receive 1 credit point dispensation for part of the financial accounting course if they followed a specific course during secondary education.

experiment and informed students that participants would be randomly assigned to three equally sized groups with equal odds for all students. Furthermore the letter explained that participation implied that the student granted the researchers permission to link information from the experiment to information from the student records about their achievements. Students received a fixed payment of NLG 50 (€ 22.69) for returning a completed participation form including a small questionnaire. After a reminder and a telephone round, 249 eligible students participated in the experiment which is 98% of all eligible students. Three students could not be reached and 2 students explicitly rejected participation. The short questionnaire collected information on respondents' math grades in high school and their parents' education.

Participants were randomly assigned to three different groups: a control group, a large reward group and a small reward group. To earn the reward treated students had to pass all first year requirement before the start of the next academic year. This is: they were required to collect 60 credit points in a period of one year.

The reward sizes of the large and small reward groups are NLG 1,500 (€ 681) and NLG 500 (€ 227). Given the substantial increase in the pass rate attributed to the NLG 1,000 (€ 454) reward for econometrics students, we judged that the rewards in the present experiment were sufficiently large to increase pass rates. Moreover, calculations made at the start of the study showed that the increase in passing rates that is necessary to obtain some reasonable statistical power is well within the 0.22-increase in passing rates that was found in the earlier study among econometrics students.

The random assignment was done by stratifying the participants on the basis of their math score in high school and their parents' education.[5] This precludes that the random assignment procedure accidentally results in groups that differ in these observed characteristics. Because we will investigate differences in incentive effects between high and low ability students, similarity of the ability distribution between the different groups is crucial. In the random assignment 83 students were assigned to the large reward group, 84 students to the small reward group, and 82 students to the control group.

On November 29, letters were sent informing participants about their assignment status. The first exam of the first term was on November 28, the others in December. The exams of the second and third term took place next calendar year in March/April and June/July respectively. The re-take exams were held in August.

Table 1 presents descriptive statistics of the background characteristics of the complete sample and for the three rewards groups. Furthermore, we split the sample in a low (below average) ability group and a high (above average) ability group. Ability depends on students' high school math grades (which can range from 1 (worst) to 10 (best)). [6] The

[5]Based on their high school math score and parental education we stratified all participants in 20 groups. If the number of students in a group could be divided by 3, exactly one-third of the students was assigned to each group. If such an equal division was not possible, one or two students were first assigned randomly to the groups.

[6]The Dutch pre-university secondary education offers two programs in math: math A and math B. Math A is considerably less advanced than math B. Students are allowed to take exams in both programs,

table shows that the randomization balances the characteristics well between the treatment and control groups. This is confirmed by ranksum tests on ability that compare the control vs the low reward group (p=0.669), the control vs. the high reward group (p=0.746), and the low reward vs. the high reward group (p=0.898).

The pre-assignment questionnaire also asked participants their subjective probability of fulfilling the requirement of passing all exams within the first academic year if they would be assigned to the control group, the small reward group and the large reward group respectively. This was done to get some indication of the anticipated effect of the rewards before the experiment actually took place. The average expected probabilities are reported in the bottom part of Table 1. Without a reward the expected pass rate equals 0.55. Given that the actual pass rates in previous years were around 0.20, students seem overconfident at the beginning of their study. If students would be entitled to the small reward the expected pass rate increases to 0.63, and it increases to 0.71 for the large reward. This implies that ex-ante the students themselves expected quite sizable effects from the rewards. No differences are observed across groups. Conditional on ability, proxied by the available math grades, the self-assessed pass probability for the control treatment could be interpreted as a measure of intrinsic motivation. We will add this as a control variable in the analyses.

After the experiment ended a second questionnaire was sent to all participants. Upon completion, students received a payment of €25. In total 234 participants responded, which is 94% of all participants. This post-experiment questionnaire asked questions concerning the time students spent on their studies during the past year, their work activities during the past study year, and possible supplementary rewards offered by third parties.

4 Results

4.1 Achievement

We report the impact estimates on four different outcome variables. First we look at the first year pass rate since the bonus was tied to collecting all 60 credit points. Second we consider the number of credit points students collected in the first year. We also look at the longer term impacts of the financial rewards, namely the number of credit points after three years, and whether the student dropped out by this time (i.e. is no longer registered as an economics undergraduate student).

Table 2 shows the sample means for all outcome variables for all students together and by treatment and ability groups. The pass rate increases with the size of the reward from

but it is not compulsory to do math A in order to do math B. The better students enroll in math B and often (40 percent) take math A on the side. Of those who do, the math A grade is on average 1.5 points higher than the math B grade. A student is assigned to the high ability group if the individual either scored a six or higher for the math B exam or an eight or higher for the math A exam. Otherwise the student is assigned to the low ability group. This results in 107 students in the high ability group and 142 students in the low ability group. Splitting the sample exactly in two is not possible because of the discreteness of the math grades.

Table 1: Descriptive Statistics

		Full sample			High ability			Low ability		
	All	Control	Small	Large	Control	Small	Large	Control	Small	Large
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Schooling father (years)	13.4	13.4	13.5	13.4	14.2	14.2	14.4	12.7	13.0	12.7
	(3.4)	(3.4)	(3.3)	(3.6)	(3.4)	(2.5)	(2.8)	(3.3)	(3.7)	(4.0)
Schooling mother (years)	12.3	12.3	12.1	12.6	12.6	13.2	13.5	12.0	11.4	12.0
	(3.1)	(3.0)	(3.3)	(3.0)	(2.9)	(2.8)	(2.6)	(3.1)	(3.5)	(3.1)
No math A	0.24	0.26	0.23	0.23	0.46	0.38	0.38	0.07	0.12	0.12
	(0.43)	(0.44)	(0.42)	(0.42)	(0.51)	(0.49)	(0.49)	(0.26)	(0.33)	(0.33)
Grade math A	6.9	7.0	6.8	7.0	8.2	7.9	8.0	6.3	6.4	6.5
	(1.1)	(1.2)	(1.0)	(1.0)	(0.9)	(0.9)	(0.7)	(0.8)	(0.7)	(0.7)
No math B	0.61	0.61	0.61	0.63	0.28	0.29	0.24	0.91	0.82	0.90
	(0.49)	(0.49)	(0.49)	(0.49)	(0.46)	(0.46)	(0.43)	(0.29)	(0.39)	(0.31)
Grade math B	6.4	6.4	6.4	6.3	6.7	7.0	6.6	4.5	4.8	5.0
	(1.2)	(1.3)	(1.2)	(1.1)	(1.1)	(0.9)	(1.1)	(0.6)	(0.4)	(0.0)
Self-assessed pass probability										
-Without reward	0.55	0.55	0.53	0.57	0.64	0.57	0.65	0.47	0.50	0.52
	(0.25)	(0.27)	(0.25)	(0.25)	(0.25)	(0.27)	(0.21)	(0.25)	(0.23)	(0.26)
-With small reward	0.63	0.63	0.60	0.65	0.71	0.67	0.72	0.56	0.55	0.60
	(0.23)	(0.24)	(0.23)	(0.22)	(0.20)	(0.22)	(0.18)	(0.24)	(0.23)	(0.23)
-With large reward	0.71	0.71	0.69	0.74	0.78	0.77	0.79	0.64	0.63	0.71
	(0.22)	(0.22)	(0.23)	(0.21)	(0.17)	(0.20)	(0.16)	(0.24)	(0.23)	(0.24)
Sample size	249	82	84	83	39	34	34	43	50	49

Note: Sample means, and standard deviations in parentheses.

Table 2: Descriptive statistics achievement, by reward and ability groups

		Full sample			High ability			Low ability		
	All	Control	Small	Large	Control	Small	Large	Control	Small	Large
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pass rate, year 1	0.213	0.195	0.202	0.241	0.333	0.382	0.441	0.070	0.080	0.102
	(0.026)	(0.044)	(0.044)	(0.047)	(0.076)	(0.085)	(0.086)	(0.039)	(0.039)	(0.044)
Credit points, year 1	32.5	33.2	31.6	32.7	39.7	39.8	46.4	27.2	26.1	23.2
	(1.4)	(2.4)	(2.4)	(2.5)	(3.3)	(3.7)	(2.8)	(3.1)	(3.0)	(3.1)
Credit points, year 1-3	83.1	84.3	81.8	83.1	94.5	100.0	118.7	75.0	69.5	58.4
	(4.0)	(6.7)	(6.9)	(7.2)	(9.7)	(10.6)	(9.7)	(9.1)	(8.7)	(8.5)
Drop-out rate	0.365	0.402	0.345	0.349	0.359	0.235	0.118	0.442	0.420	0.510
	(0.031)	(0.031)	(0.052)	(0.053)	(0.078)	(0.073)	(0.056)	(0.077)	(0.071)	(0.072)

Note: Sample means, and standard errors in parentheses.

0.195 in the control group to 0.202 and 0.241 in the small and large reward groups. This increasing pattern is present in both the high and the low ability group. The differences are, however, not significant. Notice also the large difference in pass rates between high and low students. High ability students in the control group have much higher pass rates than low ability students in the large reward group. This difference is significant (p<0.01).

At the end of the first year, students have on average collected 32.5 credit points. The differences between the control and reward groups are small and not significant. After three years, the differences are again small and insignificant. These findings, which are averages for the full sample, hide opposing effects for the high and low ability groups. The numbers of credit points after one and three years in the high ability group is increasing with the reward, while the opposite holds in the low ability group.

After three years, the drop-out rate is about 36.5 percent. Students in the reward groups are more likely to be still enrolled after three years than students in the control group. It is interesting to see that the drop-out rate is lowest for high ability students in the large reward group, while the drop-out rate is highest for low ability students in the large reward group.

Figure 1 shows separately for the high and the low ability group the cumulative distribution functions of the number of credit points achieved after one and three years for the three groups. The two top panels show the outcome distributions for high ability students, whereas the two bottom panels show them for low ability students. It is immediately clear from the graphs that for high ability students the outcome distribution for the large reward group stochastically dominates the distributions of the other two groups. For the low ability students we observe the opposite, the distribution for the large reward group is stochastically dominated by the outcome distribution for the small reward and control groups. The large reward therefore seems to have a positive incentive effect at the top of the ability distribution, and a negative effect at the bottom of the ability distribution. The ordering of the distributions seems to be monotonic with reward size, with the order reversed at the bottom of the ability distribution. This is especially clear after three years.[7]

The descriptive statistics in Table 2 and the graphs in Figure 1point in the direction of strong interaction effects between reward size and ability. We now test this more formally in a regression framework in which we estimate linear regressions of the form

$$y_i = \alpha + \delta_S D_i^S + \delta_S^A D_i^S A_i + \delta_L D_i^L + \delta_L^A D_i^L A_i + \gamma A_i + x_i' \beta + \varepsilon_i \qquad (1)$$

where D_i^S and D_i^L indicates if student i is assigned to the small and large reward group, respectively. The variable A_i is the ability index based on the high school math score of student i, measured in a scale from 1 to 10. As additional regressors we include parental education and the student's self-assessed pass probability (if the student would not have promised a bonus). We include these controls to correct for remaining differences between

[7] We also calculated two-sample Wilcoxon rank-sum (Mann-Whitney) tests. For the low ability group we find that with probability 0.165 the credit point distribution after three years is the same for the control group and the large reward group. For the high ability group the p-value on the test is 0.146.

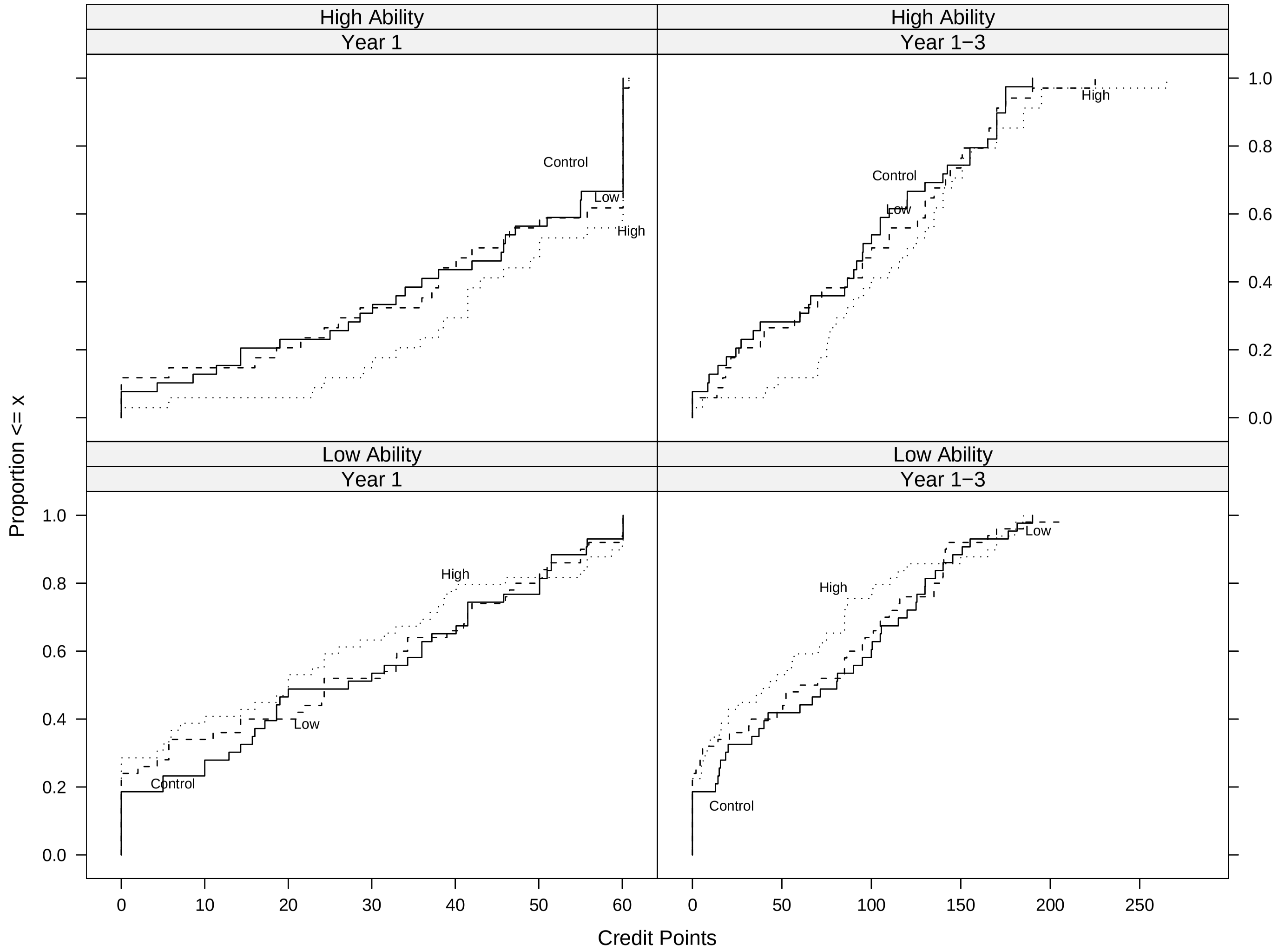

Figure 1: Cumulative distributions of credit points by treatment status for low and high ability students

groups and to reduce the residual variation in order to improve the precision of our effect estimates.

We estimate the equation both with and without the interaction term between the reward size and the ability index. If we ignore these interaction terms, the parameters δ_S and δ_L give the average treatment effects of promising respectively a small and a large reward. Once we add the interaction terms, the model allows for heterogeneous treatment effects. To illustrate the interpretation of the coefficients with this specification, δ_S gives the main (intercept) effect of the small reward, and δ_S^A traces the effect of the small reward for different levels of ability (1 point on a ten point scale). The effect of a small reward for a student who scored a 10 (the highest possible score) on his high school math exam is then $\delta_S + 10\delta_S^A$, whereas the effect for a low ability student who scored a 4 equals $\delta_S + 4\delta_S^A$.

Table 3 shows the estimation results. First, we focus on the average treatment effects, ignoring the interaction effects between the reward size and ability. The estimate for the effect of the small reward on the pass rate is 0.014, and 0.049 for the large reward. The pass rate thus increases with reward size, but the estimates are not significant. Both after one year and after three years, there is no significant effect on the reward on the number of collected credit points. The effect of the small reward is slightly negative on both outcome measures, while the effect of the large reward is in both cases almost 0. Students in both reward groups have slightly lower drop-out rates, but again the effects are insignificant.

Next we consider the interaction effects of reward size and ability in columns (2), (4), (6) and (8). These results confirm the descriptive results in Table 2 and the graphs in Figure 1. The interaction between reward and ability has a positive effect on all outcome variables (negative for the drop-out rate) and are larger for the large reward than for the small reward.

To facilitate the interpretation, Table 4 reports impact estimates for students of mean ability, low ability (mean ability minus one standard deviation), and high ability (mean plus one standard deviation), based on the regressions in Table 3. While negligible for the mean student, the point estimates are negative for low ability students and positive for high ability students. This is observed for both the small and the large reward, moreover, impact estimates increase with reward size.

Column (1) shows the results for the first year pass rate. Effects are negative at the lower end of the skill distribution and positive at the higher end. Although the effects are not significant, it is clear that if there are any effects then these are observed for the high ability students. Column (2) shows the impact estimates on the number of credit points students achieved by the end of the first year. Here we see that high ability students in the large reward group perform significantly better than high ability students in the control group. Low ability students in the large reward group score less well than their counterparts in the control group. This effect is significant at the 5 percent level.

Columns (3) and (4) report the long term effects of the rewards. For low ability students the negative effects of the rewards after one year have been magnified. After three years, low ability students assigned to the large reward group do significantly worse than low

Table 3: Achievement

	Pass rate, year 1		Credit points, 1st Year		Credit points, Year 1-3		Drop-out	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Small reward	0.014	-0.297	-0.929	-17.719	-2.189	-64.618	-0.066	0.218
	(0.056)	(0.188)	(3.029)	(10.978)	(8.962)	(34.505)	(0.074)	(0.284)
Small reward * Ability		0.057		3.072		11.433		-0.052
		(0.039)		(1.943)		(6.244)		(0.049)
Large reward	0.049	-0.327	-0.065	-32.798	0.087	-108.259	-0.056	0.668
	(0.059)	(0.203)	(2.941)	(10.155)	(8.959)	(32.814)	(0.073)	(0.277)
Large reward * Ability		0.069		6.032		19.960		-0.133
		(0.042)		(1.764)		(5.874)		(0.046)
Ability	0.120	0.082	6.704	4.006	16.089	6.728	-0.078	-0.023
	(0.017)	(0.027)	(0.883)	(1.343)	(2.785)	(4.395)	(0.022)	(0.036)
Schooling father	0.002	0.003	0.101	0.152	0.895	1.094	-0.002	-0.002
	(0.009)	(0.009)	(0.491)	(0.485)	(1.441)	(1.437)	(0.012)	(0.012)
Schooling mother	-0.005	-0.008	-1.049	-1.189	-4.040	-4.547	0.014	0.016
	(0.009)	(0.009)	(0.559)	(0.563)	(1.627)	(1.620)	(0.013)	(0.013)
Self-assessed pass prob.	0.247	0.226	21.334	22.440	43.538	47.551	-0.323	-0.343
	(0.085)	(0.084)	(4.939)	(4.767)	(15.157)	(14.444)	(0.124)	(0.121)
Intercept	-0.559	-0.345	-3.801	11.409	9.803	62.415	0.855	0.548
	(0.123)	(0.162)	(6.705)	(8.904)	(21.860)	(28.758)	(0.164)	(0.229)

Note: Regression estimates with their standard errors in parentheses. Standard errors are heteroscedasticity robust when outcome variables are binary.

Table 4: Achievement; Reward Effects by Ability

	1st Year		Year 1-3	
	Pass rate	Credit points	Credit points	Drop-out
	(1)	(2)	(3)	(4)
	A. Large reward			
Mean ability - S.D.	-0.046	-8.331	-27.301	0.127
	(0.057)	(3.869)	(12.677)	(0.107)
Mean ability	0.048	-0.124	-0.149	-0.055
	(0.059)	(2.886)	(8.736)	(0.072)
Mean ability + S.D.	0.142	8.082	27.004	-0.237
	(0.100)	(3.634)	(11.963)	(0.083)
	B. Small reward			
Mean ability - S.D.	-0.066	-5.258	-18.242	0.008
	(0.052)	(4.025)	(12.383)	(0.106)
Mean ability	0.012	-1.079	-2.688	-0.062
	(0.055)	(3.005)	(8.844)	(0.074)
Mean ability + S.D.	0.089	3.100	12.866	-0.133
	(0.096)	(3.979)	(12.582)	(0.092)

Note: Based on regressions in Table 4. Standard errors in parentheses.

ability students assigned to the control group. This group also has a higher drop-out rate (although this lacks precision). For low ability students assigned to the small reward group the effects after three years are smaller (and insignificant) but go in the same direction. For the high ability students the effects obtained after one year have also increased after three years. High ability students assigned to the large reward group collected more credit points than high ability students assigned to the control group. This effect is very substantial: 27 credit points is equivalent to half a year of nominal study duration. They also have a lower drop-out rate. Both differences are significant. For high ability students assigned to the small reward group effects are smaller and lack precision, but go in the same direction.

We thus find that there is significant heterogeneity in the behavioral response to financial incentives. Financial incentives cause low ability students to perform worse, and high ability students to perform better. Conditional on ability these relationships are monotonic over the range of the rewards that were offered. These results are consistent with the framework proposed by Camerer and Hogarth (1999) where the introduction of financial rewards causes two effects: a loss of intrinsic motivation and an incentive effect. Whether the incentive effect is sufficiently large to exceed the loss of intrinsic motivation depends on the gap between unincentivized achievement and the achievement threshold to which the incentive is tied. If for low ability students this gap is large and the threshold infeasible this implies that their incentive effect will be zero so that the loss of intrinsic motivation dominates. For students at the higher end of the skill distribution the threshold is feasible. If the gap is positive a positive incentive effect may dominate the loss of intrinsic motivation. This is consistent with results reported in Angrist and Lavy (2005) who find that only the girls in the upper part of the ability distribution respond to the rewards offered in their experiment.

Interestingly, while the experiment only rewarded achievement in the first year, the effects after three years are larger than the effects observed after one year. An explanation for the increased negative effect of the rewards on the performance of low ability students is that the poor performance and lack of positive feedback during the first year accumulates the student's discouragement. While the high ability students in the reward groups experience in the first year that working hard resulted in good exam scores, and that this motivated them to continue working hard after the experiment finished. An alternative explanation for the amplification of the (negative and positive) effects is that being on track after the first year makes studying in subsequent years easier for at least two reasons. First, making re-take exams distracts attention from the regular program. Students who have to do few or no re-take exams can follow the standard program and concentrate their attention on fewer courses. Second, students who fail exams may consequently lack the prerequisite knowledge to successfully complete second (and third) year courses.

4.2 Effort and time allocation

The effect of the rewards on achievement is a reduced form effect. It does not disentangle the effects of rewards on effort and subsequently of effort on achievement. Furthermore, promising rewards might change study behavior. It may, for example, be the case that (low ability) students in the control group focus on only a limited number of exams, while students in the reward groups have to do all exams to make a chance of earning the reward. This might cause that students in the reward groups have a lower probability of passing a particular exam than students in the control group. We collected information about students' study behavior and effort levels to examine whether the rewards had an impact on effort. The administrative records provide the number of exams taken by each student and the post-experiment questionnaire included the following questions:

- "How many hours per week did you on average spend on your study in economics and business during each of the three trimesters of the past academic year ? (We want to know the total average time spent on your study, this means including following and preparing lectures and courses and preparing exams.)"
- "How many hours did you spent in total on preparing re-take exams held in August? (Here we want to know the total number of hours, not the average per week.)"

Table 5 reports descriptive statistics on the various effort and time allocation measures by reward and ability groups. This shows that there are no substantial differences in exam taking between the control group and the reward groups. High ability students take about 9 regular exams and 2 re-take exams, while low ability students take about 7 regular exams and 2.5 re-take exams. Among the low ability students, the students in the control group

Table 5: Descriptive statistics time allocation, by reward and ability groups

		Full sample			High ability			Low ability		
	All	Control	Small	Large	Control	Small	Large	Control	Small	Large
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Exams taken, 1st Year	7.86	8.12	7.56	7.87	8.90	8.53	9.53	7.51	6.90	6.71
	(0.24)	(0.42)	(0.43)	(0.42)	(0.52)	(0.55)	(0.39)	(0.60)	(0.60)	(0.61)
Re-take exams taken, 1st Year	2.27	2.30	2.08	2.41	1.87	1.82	2.26	2.70	2.26	2.51
	(0.14)	(0.25)	(0.22)	(0.28)	(0.24)	(0.37)	(0.34)	(0.41)	(0.28)	(0.41)
Weekly study hours, 1st term	22.2	23.7	21.8	21.1	22.5	21.8	22.1	24.7	21.8	20.3
	(0.7)	(1.2)	(1.2)	(1.4)	(1.6)	(1.7)	(1.9)	(1.8)	(1.6)	(1.9)
Weekly study hours, 2nd term	18.2	18.9	17.7	18.2	19.1	18.9	19.3	18.7	17.0	17.4
	(0.8)	(1.3)	(1.3)	(1.5)	(1.6)	(1.8)	(1.9)	(2.0)	(1.9)	(2.1)
Weekly study hours, 3rd term	16.6	16.8	16.9	16.1	17.2	17.2	17.6	16.5	16.7	15.0
	(0.8)	(1.3)	(1.4)	(1.5)	(1.6)	(1.8)	(2.0)	(2.0)	(2.0)	(2.1)
Total study hours, summer	27.4	29.5	22.5	20.5	25.0	20.6	22.6	33.4	23.8	36.2
	(2.3)	(4.3)	(3.5)	(4.4)	(4.5)	(4.4)	(4.2)	(7.0)	(5.2)	(6.7)
Has job	0.80	0.76	0.83	0.81	0.77	0.88	0.78	0.75	0.79	0.82
	(0.03)	(0.05)	(0.04)	(0.05)	(0.07)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)
Weekly hours in job	12.1	12.5	12.4	11.5	12.5	13.8	9.5	12.5	11.2	12.8
	(0.5)	(0.5)	(0.8)	(1.0)	(1.6)	(1.2)	(0.9)	(1.2)	(1.1)	(1.4)
Hourly wage in job	7.6	7.7	7.9	7.2	7.6	8.0	7.1	7.8	7.8	7.3
	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.5)	(0.7)	(0.5)	(0.4)	(0.3)
Joined fraternity	0.26	0.23	0.26	0.29	0.26	0.32	0.34	0.20	0.21	0.24
	(0.03)	(0.05)	(0.05)	(0.05)	(0.07)	(0.08)	(0.09)	(0.08)	(0.06)	(0.06)
Left parental home	0.52	0.48	0.59	0.49	0.57	0.65	0.56	0.40	0.54	0.44
	(0.03)	(0.06)	(0.05)	(0.06)	(0.08)	(0.08)	(0.09)	(0.08)	(0.07)	(0.07)

Note: Sample means, and standard errors in parentheses.

take most exams. This rules out the possible explanation for the negative effect of rewards on the achievement of low ability students that the reward forces these students to take too many exams with a subsequent high failure rate. The finding that low ability student take fewer exams when being assigned to a treatment group is another indication that their intrinsic motivation is reduced due to a reward. For high ability students, those in the large reward group are slightly more likely to do re-take exams during the summer.

In the full sample, average study time in the control group is 23.7 hours per week during the first trimester and decreases to about 19 during the second trimester and 17 during the third trimester. Students in the control group spend on average 29.5 hours to prepare their re-take exams during the summer. Quite a few students report that they do not spend time at all on their study, which influences the averages for the second and third trimesters and for the summer period. These are the students who dropped out and, for the summer period, also students who did no re-take exams.[8] Students in the treatment groups tend to spend slightly less time on their study, but average time spent on the study is very similar across groups. Differences across groups are neither substantial nor significant.

In the high ability group there are hardly any differences in study effort between the students in the control group and both reward groups. In the low ability group the students in the high reward group devote on average less time to studying than students in the control group.

The questions about study time measure actual effort only imperfectly. The responses are subjective and retrospective, and only measure time input and not the effective input per hour. While biases due to this may cancel out in across group comparisons, it is desirable to have additional information about study effort. The questionnaire therefore also included items concerning time spent on paid work, whether respondents joined a fraternity and whether they lived with their parents.

About 80 percent of the students combine studying with work, and those who work spend slightly more than 12 hours per week on this activity while earning on average €7.6 per hour. Here, we see no differences between the reward and control groups with the exception that high ability students in the large reward group work less than all other groups. Finally, the rewards did not prevent students from joining a fraternity or from moving out of their parents' house.

Next, we estimate regression equations for study effort and exam taking, which are the most relevant effort and study behavior variables. In these analyses we use average weekly study effort, which is constructed using the weekly study effort in each of the three terms and the effort during the summer period. Each term consists of 14 weeks and therefore total annual study effort is 14 times the weekly study effort in each of three terms plus total study effort during the summer. Since the re-take exams take place during two weeks in the summer, the total academic year consists of 44 weeks. Therefore, the average weekly

[8]In the first trimester 3 respondents report zero study effort, in the second trimester this equals 33 and in the third trimester 39; 83 students spent zero hours on preparing for the August re-take exams, of which 22 students did not have to do any re-take exams. For the sample reporting positive numbers, the distribution of study time is bell-shaped.

Table 6: Time allocation

	Total effort		Exams taken	
	(1)	(2)	(3)	(4)
Small reward	-1.07	-5.03	-0.80	-6.29
	(1.55)	(6.57)	(0.80)	(3.45)
Small reward * Ability		0.73		1.01
		(1.13)		(0.57)
Large reward	-0.98	-5.19	-0.16	-7.52
	(1.61)	(6.52)	(0.79)	(3.47)
Large reward * Ability		0.77		1.36
		(1.13)		(0.57)
Ability	0.30	-0.18	0.51	-0.20
	(0.54)	(0.77)	(0.27)	(0.42)
Schooling father	0.15	0.16	0.12	0.14
	(0.28)	(0.28)	(0.12)	(0.13)
Schooling mother	-0.59	-0.61	-0.39	-0.43
	(0.28)	(0.29)	(0.15)	(0.15)
Self-assessed pass prob.	8.16	8.47	3.15	3.49
	(2.84)	(2.84)	(1.40)	(1.39)
Intercept	18.54	21.17	-9.08	13.06
	(3.35)	(4.47)	(1.92)	(2.67)

Note: Regression estimates with their standard errors in parentheses. Standard errors are heteroscedasticity robust when outcome variables are binary.

study effort is the total annual study effort divided by 44. Furthermore, we use the total number of exams taken, which is the sum of the regular exams and the re-take exams.[9]

Table 6 shows the estimation results. Although for self-reported effort none of the effects is significant, the estimates are consistent with the results on achievement: low ability students in the reward groups spent less time studying, whereas high ability students in the reward groups report that they spent more time on their study. The fact that we do not find significant effects of the rewards on students' self-reported effort is likely to be at least partly due to measurement error in the effort variables. This is also what Kremer et al. (2004) report. They find significantly positive effects of their rewards on observed school attendance, while the effects of rewards on self-reported measures of effort (and attitudes towards education) are insignificant.

On average students in the reward groups make slightly fewer exams than students in the control group, but these differences are not significant. However, if we consider again the ability interaction, the effects show the same pattern as reported earlier. High ability students make more exams when being assigned to a reward group, while low ability students make fewer exams when being assigned to one of the reward groups. These interactions effects are significant al the 10 percent level for the small reward and at the

[9]It should be noted that 22 students pass all regular exams and thus do not make re-take exams.

Table 7: Incidence and size of supplementary rewards

	Incidence rate (1)	Mean reward size (2)
Large reward	0.104	€770
Small reward	0.025	€750
Control	0.053	€625

5 percent level for the large reward. This again confirms that the reward stimulates the study behavior of high ability students, while reducing the intrinsic motivation of low ability students.

5 Threats to validity

Three confounding factors may potentially threat the validity of our findings.. First there may be treatment substitution bias. Parents may promise a reward or supplement the reward if students are assigned to the control or small reward group. In this case all participants would be confronted with essentially the same treatment and we would find no difference between the original three groups. To investigate whether such responses actually took place, we included in the post-experiment questionnaire a question whether someone else (for instance parents) promised a reward for passing all first year exams. Table 7 reports for each group the shares of students responding affirmative to this question along with the mean values of the size of these supplementary rewards. The table shows that supplementary rewards are fairly uncommon, and that incidence rate and size of such rewards are higher among the large reward group than among the small reward group and the control group. Therefore we expect supplementary rewards to have no impact on our findings. Note that the negative treatment effect for low ability students can only be explained by substitution bias if the low ability students in the control group were promised rewards exceeding the rewards of the experiment. Such a pattern is clearly not present in the data.

A second possible confounding factor is that teachers may grade exams differently for students in the reward groups than for students in the control groups. Although teachers are in principle unaware of the treatment status of their students, students could communicate their status in the hope that teachers will grade their exams more favorably. This seems unlikely for two reasons. First, students from the control group could also claim that they belong to a reward group if this implies that their exam will be graded more favorably. A second and more important reason is that during the first academic year most exams are multiple-choice tests. Such tests give teachers little leeway to manipulate grades of particular students, so that it can neither explain the positive treatment effect for high ability students nor the negative treatment effects for low ability students.

A final possible confounding factor is that if the rewards induce students in the reward groups to work harder, this could spill over to their peers in the control group. We consider it unlikely that spillover effects influenced our findings. The overall pass rate of the students in our experiment, and in particular the control group, is very similar to the pass rates of previous cohorts. Information about student effort from previous cohorts is in line with student effort among the students that participated in the experiment. There is also no change in the composition of the student population in terms of secondary school grades for mathematics. To attribute the negative treatment effect for low ability students together with the positive treatment effect for high ability to spillovers, it should be the case that low ability controls benefit more from spillovers than high ability controls, which is a very unlikely scenario.

6 Conclusion

This paper reports about a randomized experiment that investigated the effects of financial incentives on undergraduate students' achievement. The target population consists of first year economics and business students at the University of Amsterdam. The students, who were randomized into three reward groups, could earn a reward upon passing all first year exams before the start of their second academic year. In the large reward group the reward was €681 and in the small reward group the reward was €227. Students in the control group could not earn a reward.

We find that the rewards have small and non-significant effects on the first year pass rate. There are no effects on the number of achieved credit points by the end of the first year. Further breakdown of these results shows that there is significant heterogeneity in the behavioral response to these financial incentives. High ability students have higher pass rates and collect significantly more credit points when assigned to larger reward groups. Low ability students on the other hand appear to achieve less when assigned to larger reward groups. After the first year these effects are only significant for the high ability group, but after three years the size of the effects has increased and are statistically significant for both the low and high ability group.

One interpretation of our findings follows Camerer and Hogarth (1999) and emphasizes the importance of the match between the ability of the student and the performance threshold, and how effort translates in achievement. The performance threshold tied to the reward can result in a binding participation constraint at the bottom of the ability distribution. This will then result in zero incentive effects for low ability students. If, at the same time, financial rewards have important displacement effects on intrinsic motivation this can explain a negative relationship between reward (size) and achievement for low ability students (for whom the displacement effect dominates the incentive effect), and a positive relation for high ability students (where the incentive effect dominates the displacement effect).

Our findings add to the small, mainly non-economic, literature that points to the potentially detrimental effects of financial incentives. In environments - such as education - where intrinsic motivation is important, this calls for a careful design of reward schemes.

References

Angrist, J. A., Lang, D., and Oreopoulos, P. (2006). Lead them to water and pay them to drink: An experiment with services and incentives for college achievement. Working Paper No. W12790, NBER.

Angrist, J. D., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Columbia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558.

Angrist, J. D. and Lavy, V. (2002). The effect of high school matriculation awards: Evidence from randomized trials. Working Paper Series 9389, NBER.

Angrist, J. D. and Lavy, V. (2005). The effect of high stakes on high school achievement rewards: Evidence from a group randomized trial. Unpublished manuscript.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19:7–42.

Dearden, L., Emmerson, C., Frayne, C., Goodman, A., Ichimura, H., and Meghir, C. (2001). Education maintenance allowances: The first year - a quantitative evaluation. Research Report RR257, DfES.

Dearden, L., Emmerson, C., Frayne, C., and Meghir, C. (2002). Education maintenance allowances: The first two years - a quantitative evaluation. Research Report RR352, DfES.

Field, E. (2006). Educational debt burden and career choice: Evidence from a financial aid experiment at nyu law school. Working Paper No. W12282, NBER.

Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2006). College cost and time to obtain a degree: Evidence from tuition discontinuities. Unpublished manuscript.

Hilkhuysen, P. (2000). Premie studieresultaat; experiment in de propedeuse AEO: cohort 1999/2000. Jaarverslag onderwijsinstituut 1999-2000, bijlage 7, Faculteit der Economische Wetenschappen en Econometrie, Universiteit van Amsterdam.

Kremer, M., Miguel, E., and Thornton, R. (2004). Incentives to learn. Unpublished working paper, Harvard University.

Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2003). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. CEPR Discussion Paper No. 3921.

Table A1: Overview of the first year courses in the economics and business program

	Credit points
Trimester 1 (September-December)	
- Financial accounting	5
- Microeconomics	8
- Mathematics 1	5
- Information management A	4/3
Trimester 2 (January-March)	
- Macroeconomics	8
- Management accounting	4
- Orientation fiscal economics	2
- Mathematics 2	4
- Information management B	4/3
Trimester 3 (April-June)	
- Finance	5
- Marketing	5
- Organization	5
- Statistics	5
- Information management C	4/3