The effect of financial rewards on students' achievement: Evidence from a randomized experiment[1]

Edwin Leuven Hessel Oosterbeek Bas van der Klaauw

[1]This is a revised and extended verson 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.

Abstract

In a randomized field experiment where first year university students could earn financial rewards for passing all first year requirements within one year we find small and non-significant average effects of financial incentives on the pass rate and the numbers of collected credit points. There is however evidence that high ability students collect significantly more credit points when assigned to (larger) reward groups. Low ability students collect less credit points when assigned to larger reward groups. After three years these effects have increased, suggesting dynamic spillovers. The small average effect in the population is therefore the sum of a positive effect for high ability students and a (partly) off-setting negative effect for low ability students. A negative effect of financial incentives for less able individuals is in line with research from psychology and recent economic laboratory experiments which shows 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 increase their achievement (e.g. Angrist et al. 2002; Angrist and Lavy 2002, 2005; Dearden et al. 2002; Kremer et al. 2004). This interest is in part fed by the impression that spending money to increase education inputs (e.g. computers, class size) is often a relatively ineffective way to improve student outcomes (Hanushek 1986, 1996; Hoxby 2000). Although standard economic theory predicts a positive relation between financial incentives and effort, there is little empirical evidence that shows that financial incentives are indeed an effective way to improve student outcomes.

While there is evidence that financial incentives work in many contexts, there is also a large literature in psychology and experimental economics that draws attention to potential adverse effects of incentives. 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 results: the importance of intrinsic motivation, and the match of what they refer to as "production" (the characteristics of the task at hand) and "capital" (the cognitive skills of the subjects). They conclude that (i) capital variables such as educational background, general intelligence and experience may interact with incentives, (ii) poorly capitalized individuals may perform worse, and (iii) interaction effects between capital and production may result in so-called "floor effects" (when tasks are feasible for many) or "ceiling effects" (when task are feasible for few).

The effect of financial incentives in general, and the effect of financial incentives in education in particular is therefore an empirical question. There is relatively little 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. 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 a randomized field experiment among first year undergraduate students in economics and business at the University of Amsterdam who started in the academic year 2001/2002. The experimental design was such that freshmen were randomly assigned to three 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. 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 effect of the size of the reward. This distinction is potentially important because it has been found that rewards may affect performance in a non-monotonic fashion. Gneezy and Rustichini (2000) found that where substantial rewards improve performance, small rewards may actually deteriorate achievement.

To briefly summarize our results, for the full sample we find a small and non-significant positive effect of the large reward on achievement, both measured by pass rates and numbers of collected credit points. We find evidence for the importance of the effects highlighted by Camerer and Hogarth. In particular, high ability students (those with more "capital") have higher pass rates and collect significantly more credit points when assigned to (larger) reward groups. Low ability students (those with less "capital") on the other hand appear to achieve less when assigned to the large reward group. While

[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).

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. The average treatment effect is therefore small and insignificant not because all students are unresponsive to financial incentives, but because it is an average of a positive effect at the upper end of the ability distribution and a negative effect at the lower end.

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 results for the whole population. Section 5 investigates the behavioral heterogeneity mentioned in the previous paragraph, and discusses the interpretation of these findings. Section 6 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 7 summarizes and discusses our findings.

2 Background

University education in the Netherlands is accessible to all 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 education as opposed to at the entry of higher education.

[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.

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. Oosterbeek et al. (1992) compare the labor market outcomes of graduates from the different economics and business departments in the Netherlands and find that selection corrected wage differentials are modest.

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 dept aversion is not only observed in the Netherlands, see for example Field (2006) who finds evidence for dept aversion of 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 is a 4 year program. 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.

3 Experimental design and data

3.1 Design

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., 2005) 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.

For society, 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

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

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.

In order to investigate this question using a field experiment, we randomized first year economics and business students in the academic year 2001/2002 in three groups; a control group, a large reward group and a small reward group. 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.

An important feature of our design is the distinction between a large and a small reward. This potentially allows us to distinguish between the effect of being treated as such, and the effect of the size of the reward (and investigate the presence of potential disincentive effects of small financial rewards).

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

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

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 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.

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.

3.2 Data from questionnaire

The short questionnaire, filled out at the moment of registration for participation, collected information on respondents' mathematics grades in secondary school and their parents' education.

Table 1 presents descriptive statistics of the background characteristics of the complete sample and for the three different reward groups. Parental education is reported in years. Dutch pre-university secondary education offers two programs in mathematics: mathematics A and mathematics B. Mathematics A is considerably less advanced than mathematics B. Students are allowed to take exams in both programs, but it is not compulsory to do mathematics A in order to do mathematics B. In practice the better students choose mathematics B. Apart from the average grades on the math exams, Table 1 also reports the shares of students who did not take the exam for mathematics A, or mathematics B. Exams are graded on a scale from 1 (lowest) to 10 (highest). The random assignment was done by stratifying the participants on the basis of their mathematics results and their parents' education. This precludes that the random assignment procedure accidentally results in groups that differ in these observed characteristics. Table 1 shows that the randomization balances the characteristics well between the treatment and control groups. We will control for these background variables in our analysis.

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 overly optimistic 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. We discuss the results and findings based on these data in sections 4 to 6.

4 Results

4.1 Achievement

We report the impact estimates on a number of outcome variables. First we look at the first year pass rate since the bonus was tied to collecting all 60 credit points. Moreover, we consider the number of credit points students

Table 1: Descriptive Statistics

	All	Control	Small	Large	Pr > $\|t\|$		
					(2)=(3)	(2)=(4)	(3)=(4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Schooling Father (years)	13.442	13.378	13.524	13.422	0.779	0.936	0.848
	(3.417)	(3.416)	(3.273)	(3.596)			
Schooling Mother (years)	12.333	12.293	12.119	12.590	0.726	0.527	0.336
	(3.114)	(3.041)	(3.316)	(2.992)			
No Math A	0.237	0.256	0.226	0.229	0.655	0.686	0.967
	(0.426)	(0.439)	(0.421)	(0.423)			
Grade Math A	6.932	6.967	6.846	6.984	0.549	0.932	0.449
	(1.089)	(1.211)	(1.049)	(1.016)			
No Math B	0.614	0.610	0.607	0.627	0.973	0.826	0.798
	(0.488)	(0.491)	(0.491)	(0.487)			
Grade Math B	6.375	6.438	6.364	6.323	0.814	0.707	0.891
	(1.207)	(1.268)	(1.245)	(1.137)			
Self-assessed pass probability							
-Without Reward	0.552	0.553	0.530	0.573	0.578	0.615	0.273
	(0.254)	(0.265)	(0.251)	(0.248)			
-With Small Reward	0.625	0.630	0.598	0.648	0.375	0.600	0.145
	(0.228)	(0.236)	(0.233)	(0.216)			
-With Large Reward	0.712	0.710	0.685	0.741	0.475	0.361	0.105
	(0.220)	(0.218)	(0.228)	(0.212)			
N	249	82	84	83			

Note: Sample means, and standard deviations in parentheses. p-values in columns (5)-(7) are from standard t-tests.

collected in the first year. As mentioned above, we also look at the longer term impacts of the financial rewards, namely the number of credit points achieved after 3 years, and whether the student dropped out by this time (i.e. is no longer registered as an economics undergraduate student).

For each outcome variable we first report the raw means and standard errors for every treatment (and control) group. We estimate regressions of the form

$$y_i = \alpha + \delta_L D_i^L + \delta_H D_i^H + x_i'\beta + \varepsilon_i \tag{1}$$

where D_i^L equals one if student i was in the small reward group, and zero otherwise, D_i^H is a corresponding indicator variable for the large reward group. Finally, we report estimates with and without a set of control variables x_i. We include these to correct for remaining differences between the different treatment groups and to reduce the residual variation in order to improve the precision of our effect estimates.

Column (1) in Table 2 shows the average pass rates for the different groups. The pass rate increases monotonically with the size of the reward from 0.195 in the control group to 0.241 in the large reward group. The second column shows the results from the regression without controls. The coefficients are the difference with respect to the control group which is the reference category, and p-values from standard t-tests are reported in the lower part of Table 2. Although the pass rate increases with reward size the differences are not statistically significant. It can also not be rejected that the size of the effect is the same for the small and the large reward. Furthermore, the pooled estimate (not reported) is also insignificant with a p-value of 0.368. Column (3) adds controls to the regression in the second column. The point estimates increase somewhat and the standard errors are reduced, the final estimate of the effect of the large reward 0.055 and 0.033 for the small reward. Again, these differences are not statistically significant.

Columns (4)-(6) report the results for the number of credit points by the end of the first year. Students in the control group earn on average 33.2 credit points. This number is lower in the treatment groups where students collect on average less credit points. After adding regressors the negative impact estimates attenuate somewhat and standard errors decrease, but the pattern remains the same.

Columns (7)-(9) report the total number of credit points achieved by the

Table 2: Achievement

	Pass Rate, Year 1			Credit Points, Year 1			Credit Points, Year 1-3			Drop out		
	Mean	OLS	OLS	Mean	OLS	OLS	Mean	OLS	OLS	Mean	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Control	0.195	(reference)		33.152	(reference)		84.274	(reference)		0.402	(reference)	
	(0.045)			(2.454)			(2.554)			(0.054)		
Small reward	0.202	0.007	0.033	31.621	-1.531	-0.535	81.823	-2.452	-0.053	0.345	-0.057	-0.073
	(0.045)	(0.062)	(0.055)	(2.425)	(3.450)	(2.987)	(2.524)	(9.780)	(8.965)	(0.052)	(0.075)	(0.074)
Large reward	0.241	0.046	0.055	32.716	-0.437	-0.122	83.089	-1.185	0.091	0.349	-0.053	-0.055
	(0.045)	(0.065)	(0.058)	(2.439)	(3.460)	(2.994)	(2.539)	(9.809)	(8.985)	(0.053)	(0.076)	(0.073)
Math A			0.111			5.088			14.538			-0.077
			(0.026)			(1.384)			(4.154)			(0.034)
Math B			0.140			3.412			8.679			-0.048
			(0.031)			(1.694)			(5.085)			(0.034)
No Math A			1.019			41.500			122.815			-0.609
			(0.207)			(11.174)			(33.538)			(0.275)
No Math B			0.892			12.126			42.274			-0.242
			(0.189)			(11.045)			(33.150)			(0.241)
Educ Father			0.001			0.118			0.913			-0.002
			(0.009)			(0.477)			(1.430)			(0.012)
Educ Mother			-0.009			-1.171			-4.266			0.015
			(0.009)			(0.533)			(1.601)			(0.013)
Prob Pass			0.239			21.738			47.079			-0.348
			(0.086)			(5.011)			(15.040)			(0.124)
Intercept		0.195	-1.565		33.152	-19.004		84.274	-55.890		0.402	1.259
		(0.044)	(0.267)		(2.454)	(15.132)		(6.957)	(45.416)		(0.054)	(0.361)
Test of equality ($\Pr > \lvert t \rvert$)												
Control = Small		0.907	0.551		0.658	0.860		0.802	0.995		0.449	0.326
Control = Large		0.478	0.340		0.900	0.968		0.904	0.992		0.485	0.452
Small = Large		0.551	0.688		0.751	0.890		0.897	0.987		0.955	0.800
Small = Large = Control		0.752	0.628		0.900	0.983		0.969	0.999		0.699	0.582

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

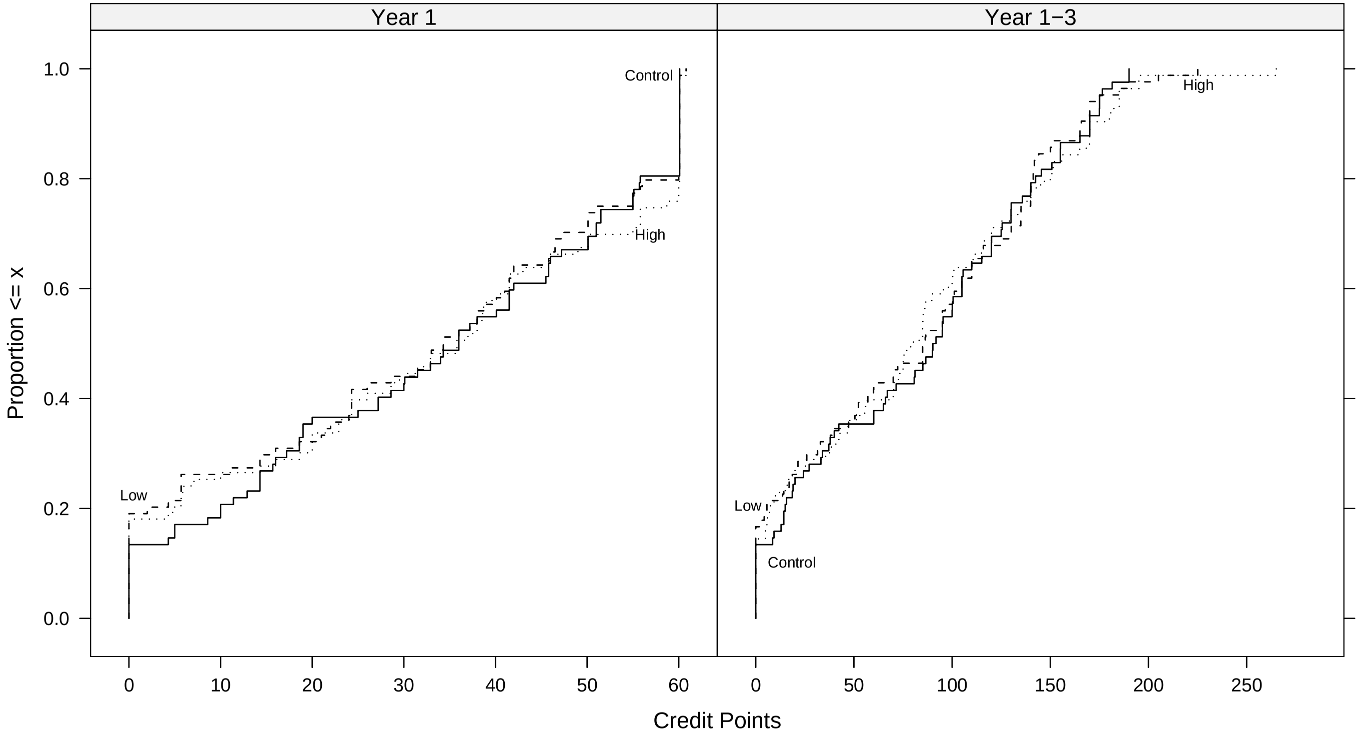

Figure 1: Cumulative distributions of credit points by treatment status

end of the third year for every treatment group. The picture here is roughly the same as after one year, small positive or negative differences between the treatment groups and the control groups, none of which is remotely significant. Finally, columns (10)-(12) report the fraction of students that dropped out of economics after three years. Students in the treatment groups are more likely to be still enrolled after three years than the students in the control group. Again these effects are not statistically significant at conventional levels.

Interestingly the point estimates on the number of credit points, both after 1 year and after 3 years, and the drop-out rate are not monotonic. Students in the small reward group perform worse than students in the control group, and students in the large reward group perform uniformly better than students in the small reward group. This is a similar pattern as in the study by Gneezy and Rustichini (2000) who find that small rewards can have disincentive effects. The fact that the average pass rates do increase monotonically with the reward size suggests that the effects of the rewards are not uniform across the credit point distribution and possibly different for different subjects.

To investigate this further, Figure 1 shows the cumulative distributions of the number of achieved credit points for each of the three reward groups after one year, and after three years. Although the three distributions are very similar after 1 year, students in the treatment groups are more often at the bottom and at the top of the credit point distribution. This explains why we observe a monotonic effect of reward size on the pass rate together with a non-monotonic (negative) impact on the number of credits points. The p-value of a chi-square test for equality of the distributions equals 0.97, indicating that by the end of the first year there are no significant differences between the distributions of credit points of the different groups. For the number of credit points collected after 3 years the cumulative density function of the treatment groups are very similar.

That students in the reward groups perform better at the top of the credit point distribution can be explained by the fact that the reward was explicitly tied to the pass rate. This is also consistent with the fact that the effect at the top is only observed in the first year, when the reward was in place, and not after three years. It is more difficult to explain the poorer performance of students in the small and large reward groups at the bottom of the distribution. One explanation could be an interaction effect between the financial reward and student ability (Camerer and Hogarth, 1999). If there are ceiling effects, then low ability students might perceive the threshold of the 60 credits points necessary to collect the reward as infeasible. The financial reward would not have an incentive effect for this group. If in addition the external reward displaces intrinsic motivation of students (Frey and Oberholzer-Gee, 1997), then financial incentives would deteriorate the performance of low ability students. High ability students on the other hand, might judge the requirement of the reward as feasible and increase their effort. We explore this further in section 5.

4.2 Effort and time allocation

The effect of rewards on achievement is a reduced form effect. It does not disentangle the effects of rewards on effort and subsequently of effort on achievement. Even if effects on achievement are zero or small, we cannot rule out that students increased their effort. We collected information about students' effort levels to examine whether the rewards had an impact on effort. 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 (2001/2002)? (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 spend 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.)"

Information on study time is provided in the first block of Table 3.

Average study time in the control group is around 23.6 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

Table 3: Time allocation

	Hours p/w				Hours	Job			Joined	Left
	1st Term	2nd Term	3rd Term	Average	Summer	Hours>0	Hours p/w	Wage p/h	Fraternity	Home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Small Reward	-1.836	-1.109	0.076	-0.957	-6.992	0.069	-0.141	0.187	0.029	-0.105
	(1.689)	(1.906)	(1.959)	(1.659)	(5.546)	(0.065)	(1.307)	(0.425)	(0.069)	(0.080)
Large Reward	-2.588	-0.659	-0.742	-1.330	1.026	0.045	-1.025	-0.471	0.059	-0.014
	(1.822)	(1.936)	(1.990)	(1.685)	(6.106)	(0.067)	(1.340)	(0.435)	(0.071)	(0.082)
Intercept	23.653	18.853	16.827	19.778	29.480	0.760	12.509	7.688	0.227	0.520
	(1.204)	(1.378)	(1.416)	(1.199)	(4.279)	(0.050)	(0.972)	(0.314)	(0.049)	(0.058)
$\Pr > \lvert t \rvert$										
Small Reward	0.278	0.561	0.969	0.565	0.209	0.287	0.914	0.659	0.669	0.188
Large Reward	0.157	0.734	0.709	0.431	0.867	0.502	0.445	0.281	0.407	0.869

Note: Regressions without controls. Standard errors in parentheses.

summer period, also students who did no re-take exams.[5] 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 statistically significant.

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. Results are also reported in Table 3.

Columns (6)-(8) show that 76 percent of the students in the control group combine studying with work, and those who work spend around 12.5 hours per week on this activity while earning on average €7.69 per hour. Here, we see no differences between the reward and control groups with the exception that students in the large reward group tend to earn somewhat lower wage rates than students in the other two groups. Finally, the last two columns of Table 3 reveal that the rewards did not prevent students from joining a fraternity or from moving out of their parents' house.[6]

To summarize, we cannot reject equality in reported study time and other time allocation variables between the treatment and control groups. This result is consistent with the first finding that rewards do not affect achievement by a significant margin, although it should be noted that it is not possible to rule out differences given the precision of the reported estimates in Table 3.[7]

[5] 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.

[6] Adding controls to the analysis, as reported in table A2, does not change the differences between the control and reward groups.

[7] If the rewards would have increased students' study time then we would have been able to estimate the causal effect of study time on achievement. Since the rewards do not change study time we cannot estimate such an effect. Regressing the pass rate on study time we find that one hour study time extra per week is associated with a one percentage point higher pass rate. Adding controls for ability, social background and the subjective pass rate does not change the size of this correlation.

5 Ability interactions

In the previous section we found indications for both non-monotonic effects of the size of the rewards, and differential impacts of the reward across the outcome distribution. There are good reasons to expect that some students will be more responsive to a reward than others because of heterogeneity in the marginal cost of effort or heterogeneity in returns to effort. Given the performance threshold attached to the reward, high ability students have to bridge a smaller gap than low ability students because the former collect more credit points in the first place. In addition, high ability students probably earn more extra credit points than low ability students for a given increase of effort. Consequently, high ability students are more likely to respond to the rewards than low ability students.

In this section we explore this further by looking at the interaction between the rewards and student ability. To do so, we construct an ability index based on students' high school math grades (which can range from 1 (worst) to 10 (best)). Students could matriculate in two types of math. As noted in section 3.2 above, type B math is considerably more advanced than type A math. 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. We therefore take ability as the maximum of the student's math B grade and math A grade minus 1.5.[8] The distribution of the ability index has a (slightly skewed) bell shape (ranging from 2 to 9 with mean 5.4 and standard deviation 1.4).

As a robustness check we also performed the analysis in this section using an alternative ability index, namely the grade of the first (math) exam of the first term (which was held before the students learned their treatment status, but after the announcement of the experiment). The drawback of this variable is that 20 percent of the students did not make this exam on the first occasion. For this reason we present the results based on the high school math grades in this section. The results of the analysis based on the subsample of students with a non-missing grade on the first exam are however very similar to the ones presented here and therefore provide additional support for our findings.

[8] We also performed a principal component analysis of the math grades and the math indicator variables. Using the first two scores from this analysis does not change the conclusions.

Before presenting our estimation results, we first look at the raw data. To stratify by ability we split the sample in two: students of high (above average) ability and students of low (below average) ability. A student is assigned to the high ability group if he scores at least six on the index which implies at least a grade sufficient to pass the math B exam (6), otherwise a student is assigned to the low ability group. This results in 100 students in the high ability group and 149 students in the low ability group.[9]

Figure 2 shows, stratified by ability group, the cumulative distribution functions of the number of credit points achieved after 1 and 3 years for the three reward groups. The two top panels show the outcome distributions for the high ability students, whereas the two bottom panels show them for the low ability students. It is immediately apparent from the graph that for the high ability students the outcome distribution for the high reward group stochastically dominates the distributions of the other two groups. For the low ability students we observe the opposite, here the distribution for the large reward group is stochastically dominated by the outcome distribution for the lower rewards 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 apparent after 3 years.[10]

We estimated the following regression to estimate how the treatment effects vary with ability:

$$y_i = \alpha + \delta_L D_i^L + \delta_L^A D_i^L A_i + \delta_H D_i^H + \delta_H^A D_i^H A_i + \gamma A_i + x_i' \beta + \varepsilon_i \qquad (2)$$

where A_i is our ability index. As above, D_i^L is a treatment indicator for the low reward group, and D_i^H is a corresponding indicator variable for the large reward group. Finally, x_i is a set of control variables, and y_i the outcome variable. To illustrate the interpretation of the coefficients with this

[9]Splitting the sample exactly in two is not possible because of the discreteness of the math grades.

[10]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 in 2004 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. These tests are of course conservative, and do not control for covariates. Regressions without controls give similar results, see table A3.

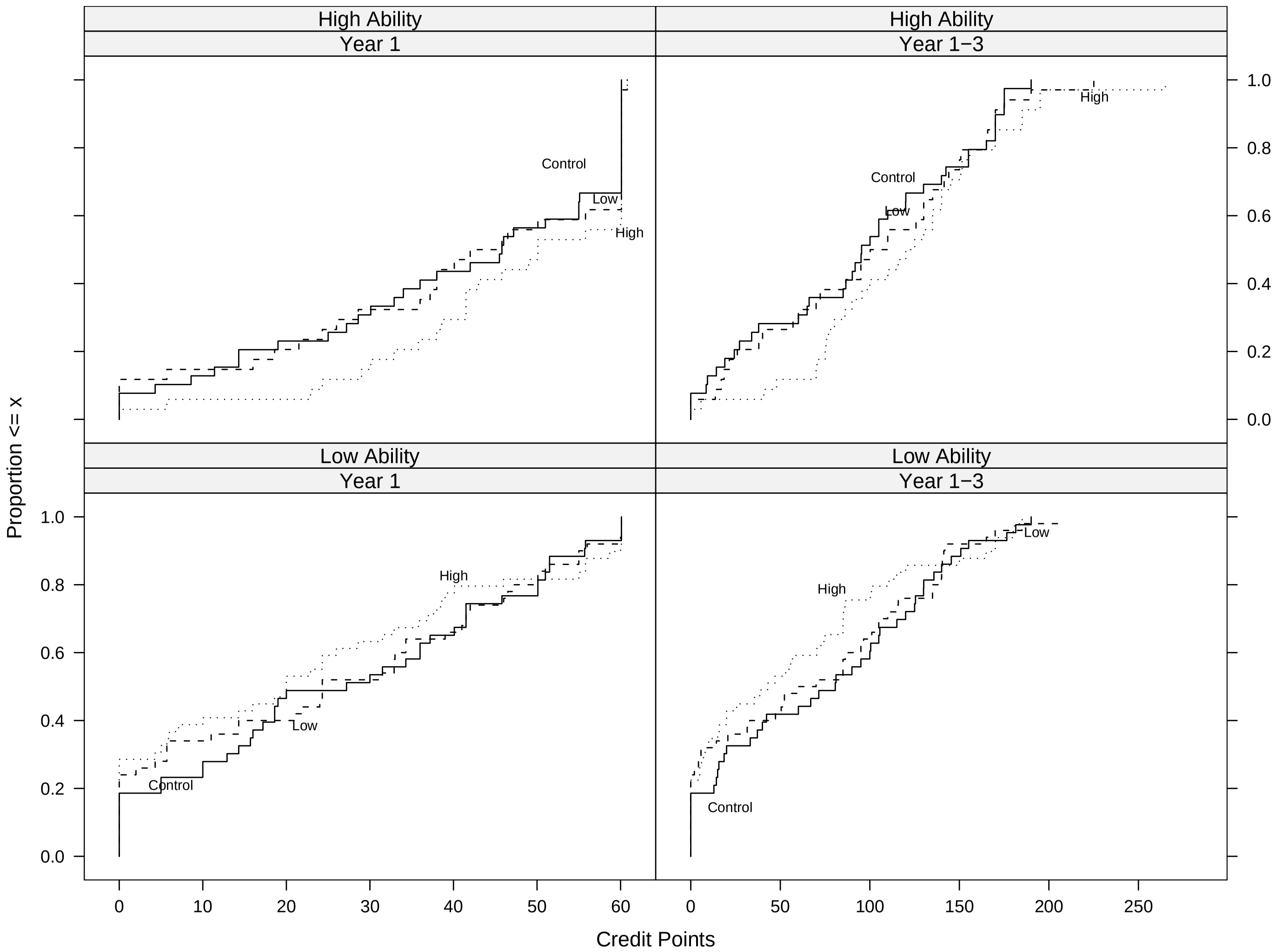

Figure 2: Cumulative distributions of credit points by treatment status for low and high ability students

Table 4: Achievement and Effort; Ability Interactions

	Achievement, Year 1		Achievement, Year 1-3		Self-reported Effort, Year 1	
	Pass Rate	Credit Points	Credit Points	Drop Out	Hours p/w	Hours Summer
	(1)	(2)	(3)	(4)	(5)	(6)
Small Reward	-0.297	-17.719	-64.618	0.218	-5.079	-7.986
	(0.188)	(12.040)	(35.650)	(0.284)	(6.611)	(20.543)
Small Reward * Ability	0.057	3.072	11.434	-0.052	0.760	0.272
	(0.039)	(2.149)	(6.364)	(0.049)	(1.185)	(3.467)
Large Reward	-0.327	-32.798	-108.259	0.668	-5.610	7.468
	(0.203)	(12.516)	(37.059)	(0.277)	(6.920)	(27.127)
Large Reward * Ability	0.069	6.033	19.960	-0.134	0.844	-1.348
	(0.042)	(2.241)	(6.635)	(0.046)	(1.239)	(4.445)
Ability	0.082	4.006	6.728	-0.024	-0.145	-1.875
	(0.027)	(1.493)	(4.421)	(0.036)	(0.854)	(2.591)
Schooling Father	0.003	0.152	1.094	-0.002	0.175	-0.445
	(0.009)	(0.483)	(1.431)	(0.012)	(0.268)	(1.144)
Schooling Mother	-0.008	-1.189	-4.547	0.017	-0.630	-0.390
	(0.009)	(0.534)	(1.582)	(0.013)	(0.295)	(0.976)
Self-assessed Pass Prob.	0.266	22.440	47.551	-0.343	8.614	10.783
	(0.084)	(4.986)	(14.762)	(0.121)	(2.754)	(8.500)
Intercept	-0.345	11.409	62.415	0.548	21.122	44.440
	(0.162)	(9.355)	(27.701)	(0.229)	(5.225)	(17.153)
Pr $> \|t\|$						
Small Reward	0.115	0.142	0.071	0.444	0.443	0.698
Small Reward * Ability	0.147	0.154	0.074	0.291	0.522	0.938
Large Reward	0.108	0.009	0.004	0.017	0.418	0.783
Large Reward * Ability	0.099	0.008	0.003	0.004	0.497	0.762

Note: Standard errors are heteroscedasticity robust when outcome variables are binary.

specification, δ_L gives the main (intercept) effect of the small reward, and δ_L^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 $\delta_L + 10\delta_L^A$, whereas the effect for a low ability student who scored a 4 equals $\delta_L + 4\delta_L^A$.

Table 4 reports how the effect estimates vary with ability. The first thing to note is that, conditional on ability, there is now a monotonic relation between the size of the reward and the impact estimate on all outcome variables. For students with a higher ability level achievement improves with reward size, whereas for students with a lower ability level achievement deteriorates with the size of the reward. The small average effects found earlier are therefore an average of negative effects at the bottom of the skill distribution, and positive effects at the top of the skill distribution. This not only explains the modest effects we found above, when negative and positive effects approximately cancel out, but is also the source of the non-monotonic relationship between reward size and credit points observed in Table 2.

Table 5 reports impact estimates for students of average (mean) ability, low ability (mean ability minus one standard deviation), and high ability (plus one standard deviation), based on the regressions in Table 4. While negligible for the average 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. This lends support for the interpretation forwarded above that there is a binding participation constraint for low ability students. 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 a 5 percent level.

Columns (3) and (4) report the long term effects of the rewards. For

Table 5: Achievement; Reward Effects by Ability

	Achievement, Year 1		Achievement, 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.

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 ability students assigned to the control group. This group also has a higher dropout 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. They also have a lower dropout rate. Both differences are statistically 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. Low ability students perform worse, and high ability students perform better. Conditional on ability these relationships are monotonic over the range of the rewards that were offered. It is possible to interpret these results along the lines of Camerer and Hogarth (1999). If financial rewards have important displacement effects on intrinsic motivation this can explain a negative relationship between reward (size) and achievement, especially at the bottom of the skill distribution. The performance threshold tied to the reward, can also result in ceiling effects at the bottom. This will then result in zero incentive effects for low ability students. For students at the higher end of the skill distribution the threshold is feasible and positive incentive effects are observed. This finding 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. The increased negative effect for low ability students could be readily explained if the aforementioned reduction of intrinsic motivation is permanent. The increased positive effect for high ability students cannot be explained by reference to intrinsic motivation. Assuming that for this group the rewards initially also have adverse effects on intrinsic motivation it would require that intrinsic motivation not only reappears but is higher than it would have been in the absence of the financial reward in the first year. One explanation for the increased positive effect for high ability students is that during the year of the experiment, the high ability students in the reward groups experienced that working hard resulted in good exam results, 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.

Are the effects on achievement consistent with self-reported effort? The last two columns of Table 4 report self-reported effort by treatment/ability interaction. Although none of these effects are significant, the point estimates are consistent with the results on achievement: low ability students in the treatment groups spent less time studying, whereas high ability students in the treatment 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-

Table 6: Incidence and size of supplementary rewards

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

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.

6 Threats to validity

While a randomized experiment is often considered the gold standard in research on treatment evaluation (Currie, 2001; Duflo and Kremer, 2003), it is not without threats to validity of the outcomes. Heckman et al. (1999) and Philipson (2000) have drawn attention to the importance of general equilibrium effects and external treatment effects or spillover effects. In the context of our experiment three confounding factors may play a role. 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 6 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.

7 Conclusion

This paper reports about a randomized social 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. In particular, 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 is statistically significant for both the low and high ability group. The average treatment effect is therefore small and non-significant not because the students are unresponsive to financial incentives, but because it is an average of a positive effect at the upper end of the skill distribution and a negative effect at the lower end.

One interpretation of our findings follows Camerer and Hogarth (1999) and emphasizes the importance of the match between the ability of the subject (capital), and how effort translates in achievement (production). The performance threshold tied to the reward can result in a binding participation constraint (i.e. ceiling effects) 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). This mixing of negative and positive relationships between reward size and achievement for different subgroups can generate non-monotonic relationships between reward size and achievement in the population as in Gneezy and Rustichini (2000).

Due to a relatively small sample, the results reported in this paper are not very precisely estimated. Yet they are clearly informative about the signs and the sizes of the effects. Because we study an entire population we can only increase the sample size by running the same experiment with new cohorts. The related studies conducted by Angrist and Lavy (2005) and Kremer et al. (2004) suffer from comparable problems with precision of the estimates. Taken together, however, the three studies reveal a consistent pattern. All three papers indicate that financial incentives for students may enhance achievement. Two out of three papers (Angrist and Lavy and the current paper) moreover report evidence that positive effects are concentrated amongst high ability students. Given the disputable effectiveness of other education interventions it seems that financial incentives for students may be a promising alternative that should be explored further.

References

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.

Currie, J. (2001). Early childhood interventions. *Journal of Economic Perspectives*, 15(2):213–238.

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.

Duflo, E. and Kremer, M. (2003). Use of randomization in the evaluation of development effectiveness. Paper prepared for the World Bank Operations

Evaluation Department (OED) conference on evaluation and development effectiveness in Washington, D.C. 15-16 july, 2003.

Field, E. (2006). Educational dept burden and career choice: evidence from a financial aid experiment at nyu law school. NBER Working Paper No. 12282.

Frey, B. S. and Oberholzer-Gee, F. (1997). The costs of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 87(4):746–755.

Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2005). Tuition discontinuities and the efficiency of educational choices. Unpublished manuscript.

Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 110:791–810.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24:1141–1177.

Hanushek, E. A. (1996). School resources and student performace. In Burtless, G., editor, *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. The Brookings Institution.

Heckman, J. J., Lochner, L., and Taber, C. (1999). General equilibrium cost benefit analysis of education and tax policies. Working Paper Series 6881, NBER.

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.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115(4).

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 randomize experiment. CEPR Discussion Paper No. 3921.

Oosterbeek, H., Groot, W., and Hartog, J. (1992). An empirical analysis of university choice and earnings. *De Economist*, 140:293–309.

Philipson, T. J. (2000). External treatment effects and program implementation bias. Technical Working Paper Series T0250, NBER.

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

Table A2: Time allocation by Reward Group

	Hours p/w				Hours	Job			Stud.	Left
	1st Term	2nd Term	3rd Term	Average	Summer	Hours>0	Hours p/w	Wage p/h	Assoc.	Home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Small Reward	-1.749	-0.950	0.293	-0.802	-6.433	0.069	0.082	0.141	0.031	-0.116
	(1.677)	(1.863)	(1.948)	(1.630)	(5.501)	(0.066)	(1.314)	(0.426)	(0.066)	(0.077)
Large Reward	-2.513	-0.465	-0.678	-1.219	1.121	0.045	-1.059	-0.481	0.056	-0.014
	(1.795)	(1.897)	(1.983)	(1.659)	(6.334)	(0.068)	(1.357)	(0.441)	(0.070)	(0.079)
Intercept	26.573	22.219	14.863	21.219	52.145	0.828	13.702	6.469	-0.336	1.827
	(8.432)	(9.412)	(9.839)	(8.233)	(25.671)	(0.339)	(6.636)	(2.140)	(0.297)	(0.354)
$\Pr > \|t\|$										
Small Reward	0.298	0.611	0.881	0.623	0.243	0.299	0.950	0.740	0.643	0.135
Large Reward	0.163	0.806	0.733	0.463	0.860	0.513	0.436	0.277	0.423	0.859

Note: Regressions with controls. Standard errors in parentheses. Standard errors are heteroscedasticity robust when outcome variables are binary. Control variables are type A math, type B math, two dummies if either is missing, education mother, education father, self-assessed pass probability.

Table A3: Achievement and self-reported effort by ability, No Controls

	Achievement, Year 1		Achievement, Year 1-3		Self-reported effort	
	Pass Rate	Credit Points	Credit Points	Drop Out	Hours p/w	Hours Summer
	(1)	(2)	(3)	(4)	(5)	(6)
Small Reward	-0.196	-7.594	-36.450	0.067	-0.320	-3.179
	(0.184)	(12.402)	(36.257)	(0.287)	(6.668)	(21.099)
Large Reward	-0.258	-25.597	-88.896	0.562	-2.343	11.679
	(0.202)	(13.048)	(38.147)	(0.287)	(7.053)	(26.907)
Small Reward * Ability	0.038	1.151	6.194	-0.023	-0.155	-0.667
	(0.039)	(2.212)	(6.468)	(0.049)	(1.192)	(3.537)
Large Reward * Ability	0.057	4.752	16.420	-0.115	0.229	-2.126
	(0.042)	(2.335)	(6.827)	(0.048)	(1.262)	(4.445)
Ability	0.100	5.305	9.382	-0.044	0.455	-1.358
	(0.027)	(1.509)	(4.411)	(0.036)	(0.838)	(2.449)
Intercept	-0.351	4.105	32.902	0.644	17.301	36.869
	(0.128)	(8.549)	(24.993)	(0.207)	(4.717)	(14.857)
Pr>|t|						
Small Reward	0.287	0.541	0.316	0.816	0.962	0.880
Large Reward	0.202	0.051	0.021	0.052	0.740	0.665
Small Reward * Ability	0.338	0.603	0.339	0.638	0.897	0.851
Large Reward * Ability	0.177	0.043	0.017	0.018	0.856	0.633

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