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

Leuven, E.; Oosterbeek, H.; van der Klaauw, B.

Published in:
Journal of the European Economic Association

DOI:
10.1162/jeea_a_00024

THE EFFECT OF FINANCIAL REWARDS ON STUDENTS' ACHIEVEMENT: EVIDENCE FROM A RANDOMIZED EXPERIMENT

Edwin Leuven
École Nationale de la Statistique
et de l' Administration Économique
CREST

Hessel Oosterbeek
University of Amsterdam

Bas van der Klaauw
VU University Amsterdam

Abstract
This paper reports on 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. (JEL: I21, I22, J24)

1. Introduction

There has recently been increased interest in the effectiveness of financial incentives for students to improve their achievement (Angrist et al. 2002; Angrist and Lavy 2002, 2009; Dearden et al. 2002; Kremer, Miguel, and Thornton 2004; Angrist, Lang, and Oreopoulos 2009). One reason for this interest is the impression that students often do not exert sufficient study effort. Standard economic theory predicts a positive relation between financial incentives and achievement. Yet insights from behavioral economics and mixed empirical evidence cast doubt on the strength of this relation in each context and for each group of individuals.

Camerer and Hogarth (1999) review 74 studies in which subjects were paid zero, small, or large financial rewards for a large variety of tasks. The effects of

The editor in charge of this paper was Orazio Attanasio.

Acknowledgments: This is a revised and extended version of Leuven et al. (2003). We thank Josh Angrist, Michael Kremer, Guy Laroque, Steve Pischke, Erik Plug, seminar participants in Amsterdam, Dublin, Sevilla, Stockholm, Tilburg, and Uppsala, and the anonymous referees for their comments.

E-mail addresses: Leuven: edwin.leuven@ensae.fr; Oosterbeek: h.oosterbeek@uva.nl; Van der Klaauw: bklaauw@feweb.vu.nl

Journal of the European Economic Association December 2010 8(6):1243–1265

incentives on performance in these studies are mixed and complicated. Camerer and Hogarth highlight two important factors to explain the variation in findings: the importance of intrinsic motivation, and the match between what is needed to earn a reward and individuals' 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. The introduction of financial incentives may have adverse effects and may (partly) crowd out students' instrinsic motivation. Students are also heterogeneous in their academic aptitudes, so 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. Ultimately the effect of financial incentives in education will therefore be an empirical question.

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 had to collect all 60 credit points in one year, where historically about 20% of students pass all first-year exams within a year and the average number of credit points collected in a year is about 30 (with a standard deviation of 22). Students assigned to the *small reward* group could earn a bonus of NLG 500 (€227) 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 potentially allows us to separate the effect of receiving a financial reward from the effect of the size of the reward. In order 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 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. In contrast, low-ability students appear to achieve less when assigned to the large reward group. At the end of the first year these effects are significant only for the high-ability group, but after three years the sizes of the effects have increased and are statistically significant for both low- and high-ability students. This suggests the presence of positive dynamic spillovers because the rewards were tied to first-year performance only. It is sometimes argued that financial rewards induce students to work harder while the incentive is in place, but not afterward. Our findings do not confirm this mechanism.

A few other studies also 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 (2009) 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 and find that the intervention led to a substantial increase in matriculation rates among girls. Angrist, Lang, and Oreopoulos (2009) evaluate how merit scholarships and study-group services affect achievement at a large Canadian university. They find no effects for boys. Girls had improved grades, which faded somewhat after a year, and the treatment that combined the merit-scholarship with peer advising and study-group services was more effective. Kremer, Miguel, and Thornton (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 the authors find 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]

The remainder of this paper is organized as follows. Section 2 provides relevant background information about the Dutch system of higher education and the economics and business program at the University of Amsterdam. Section 3 explains the design of the field experiment and describes the data and it also discusses potential threats (such as substitution bias, manipulation by teachers, and externalities) to the validity of the experiment and concludes that these are unlikely to affect our conclusions. Section 4 presents the impact of the financial rewards on performance during the first year. Section 5 discusses the evidence on students' study time and effort. In Section 6 we turn our attention to the long-term impact of the intervention on performance and dropout. Section 7 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

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 Colombia'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 covering about half of the cost of private secondary school. Vouchers were renewed only for pupils who maintained satisfactory academic performance (Angrist et al. 2002).

education qualification can be obtained only 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, so all students starting a university education in economics or business are presumed capable of actually graduating (if they exert sufficient effort). In the academic year 2001–2002 there were 34,200 first-year students at Dutch universities, which is about 17% of the relevant birth cohort. Universities are not permitted to select students; anyone who applies with a valid entry qualification must be admitted.[2] Thus, selection in the Netherlands takes place at the exit of secondary education, not at the entry of higher education.

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

The undergraduate program in economics and business at the University of Amsterdam has a nominal duration of four 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, because 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 (for students who fail the regular test) 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. (Table 1 gives an overview of the first-year courses and the number of credit points assigned to each course.) 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%. Such low pass rates are not uncommon in continental European countries (Garibaldi et al. 2007) and can be attributed to an institutional arrangement under which universities are publicly funded and tuition fees are low or nonexistent. As a consequence, students are not confronted with appropriate prices and spend more time in the system than the nominal duration of their studies. Nominal study schedules operate on the assumption that the average student should spend 40 hours per week (during 42 weeks per year) studying in order to complete in time. Surveys (including ours) asking

2. For a few programs, 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 studies in economics and business.

TABLE 1. Overview of first-year courses in the economics and business program.

Course	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

students about their actual study time typically find self-reported average amounts slightly above half that. Hence, the consensus is that the low pass-rate in the first year (and the long actual study durations) should be attributed to insufficient student effort and not to the program being too demanding. Casual observation (based on information from colleagues at different universities throughout Europe) confirms that the study program at the University of Amsterdam is not more demanding than elsewhere. This claim is also supported by comparing the results of the regular Dutch students to the results of foreign students enrolled in the English language program (which was not part of the experiment). The first-year pass rate in that otherwise similar program is more than double that of the Dutch language program.

For society, study delay imposes a cost in the form of extra expenditures on education and the forgone productivity of the students. The Department of Economics at the University of Amsterdam has an incentive to increase the pass rate because funding depends in part on the number of credit points awarded each year. There are other reasons to address the delay of students: Teaching becomes more difficult because not all students are on schedule, and the failing and re-taking of courses and 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 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, undergraduate econometrics is a separate program from economics and business, and it attracts students from the upper part of the ability distribution. In the year that this reward was in place, the pass rate was 50% compared with 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. Although the increase of 22 percentage points 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 include 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

3.1. Design

We conducted a field experiment among first-year economics and business students at the University of Amsterdam to investigate the effectiveness of financial rewards in improving achievement of university students. The experiment took place in the academic year 2001–2002.

To ensure that all students were treated identically, participation in the experiment was open only to students who (i) followed the full-time (Dutch language) program, (ii) did not claim more than 1 credit point dispensation,[3] and (iii) did not start the economics and business program in a previous year. In total, there were 254 eligible students.

Participation in the experiment was voluntary. On 1 October 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 experiment and informed students that participants would be randomly assigned to three equally sized groups with equal odds for all students. The letter also explained that participation required students to grant 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, which included a short questionnaire. After a reminder and a telephone round, 249 students participated in the experiment (98% of all eligible students). Three students

3. Students who followed a particular course during their secondary education received a 1 credit point dispensation for part of the financial accounting course.

could not be reached, and two students explicitly rejected participation. The 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 a reward, treated students had to pass all first-year requirements before the start of the next academic year. In other words, they were required to collect 60 credit points in a period of one year.

The reward sizes of the large and small reward groups were NLG 1,500 (€681) and NLG 500 (€227), respectively. Given the substantial increase in the pass rate attributed to the NLG 1,000 (€454) reward for econometrics students described in Section 2, we expected the rewards in the present experiment to be sufficiently large to increase pass rates. Moreover, calculations made at the start of the study showed that the increase in passing rates that would be necessary to obtain some reasonable statistical power was well within the 22-percentage-point increase in passing rates found in the study of econometrics students.

The random assignment was accomplished by stratifying participants in terms of their high school math score and parents' education. For math score we defined eight intervals and for parental education three intervals, resulting in $8 \times 3 = 24$ strata. Two strata (with the lowest parental education and highest math scores) did not contain any participants. The other 22 strata contained between 2 and 28 individuals. If a stratum contained more than three participants then we randomly assigned one participant to each group (high reward, low reward, and control). We repeated this process until the stratum contained fewer than three unassigned participants. Remaining unassigned participants were randomly assigned to the groups with the proviso that no two remaining unassigned participants could enter the same group. A total of 83 students were assigned to the large reward group, 84 students to the small reward group, and 82 students to the control group.

This procedure precludes random assignments that result in groups that differ in ability or parental education. Because we seek to investigate differences in incentive effects between high- and low-ability students, it is crucial that the ability distribution of the different groups is similar.

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

3.2. Data

Table 2 presents descriptive statistics of the background characteristics for the complete sample and for the three rewards groups. We also split the sample into

TABLE 2. 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, with standard deviations in parentheses.

a low (below-average) ability group and a high (above-average) ability group.[4] Ability was measured by students' high school math grades (which range, worst to best, from 1 to 10). The Dutch pre-university secondary education offers two programs in mathematics: Math A and Math B. Math A is considerably less advanced than Math B. Students are allowed to take exams in both programs, but Math A is not a prerequisite for Math B. The better students enroll in Math B and often (about 40% in our sample) take Math A on the side. For these students, their Math A grade is on average 1.5 points higher than their Math B grade. We assign to the high-ability group either students who score a 6 or higher on the Math B exam or an 8 or higher on 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 owing to the discrete nature of the mathematics grades.

The table shows that the randomization balances the characteristics well between the treatment and control groups. Only one difference between treatment groups is statistically significant at conventional levels. The difference in Math B grade between students in the control group and in the large reward group is significant with a p-value of 0.091. Note that in the low-ability subsample there are only four students in the control group and five students in the large reward group who have a grade for Math B, since most of these students took only Math A in high school. The balancing is also confirmed by rank-sum tests on ability that compare the control versus the low reward group ($p = 0.669$), the control versus the high reward group ($p = 0.746$), and the low reward versus the high reward group ($p = 0.898$).

The pre-assignment questionnaire also asked participants about their own perceived probabilities 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, or the large reward group. This gave us an 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 2. Without a reward, the average subjective anticipated pass rate is 0.55. Given that the actual pass rates in previous years were about 0.20, students seem overconfident at the beginning of their study. In the small reward scenario 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 (as proxied by the available math grades), the self-assessed pass probability for the control treatment could be

4. Students who are more likely to be credit-constrained could be more responsive to financial incentives (or students from better-off families less responsive). However, we find no indication that the incentive effects differ for students with different parental background (results not reported here). This finding is consistent with the fact that the student grant system in Netherlands is parental means-tested.

TABLE 3. 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

interpreted as a measure of intrinsic motivation. We 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, or 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.

3.3. *Threats to Validity*

Three confounding factors may threaten 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, respectively. Hence all participants would be confronted with essentially the same treatment and we would find no differences between the three groups. To investigate whether such responses actually took place, the post-experiment questionnaire asked whether someone else (for instance parents) promised a reward for passing all first-year exams. Table 3 reports, for each reward group, the share of students responding affirmatively 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 we therefore expect them to have no impact on our findings. Observe that negative treatment effects for low-ability students can be explained by substitution bias only if the low-ability students in the control group were promised rewards exceeding the rewards of the experiment. Such a pattern is 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 group. 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 doing so would mean that their exam will be graded more favorably. A second

TABLE 4. First-year achievement, by reward and ability group.

		Treatment					
	Mean	Control (1)	Low (2)	High (3)	(2) (1) (4)	(3) - (1) (5)	(3) - (2) (6)
A. Pass rate							
- All	0.213	0.195	0.202	0.241	0.007	0.046	0.039
					(0.062)	(0.065)	(0.065)
- High-ability	0.383	0.333	0.382	0.441	0.049	0.108	0.059
					(0.114)	(0.115)	(0.121)
- Low-ability	0.085	0.070	0.080	0.102	0.010	0.032	0.022
					(0.055)	(0.059)	(0.058)
DID (high- vs. low-ability)					0.039	0.076	0.037
					(0.127)	(0.129)	(0.134)
B. Credit points, first year							
- All	32.5	33.2	31.6	32.7	−1.5	−0.4	1.1
					(3.4)	(3.4)	(3.5)
- High-ability	41.8	39.7	39.8	46.4	0.1	6.7	6.6
					(5.0)	(4.4)	(4.6)
- Low-ability	25.4	27.2	26.1	23.2	−1.2	−4.0	−2.9
					(4.4)	(4.4)	(4.4)
DID (High- vs. low-ability)					1.2	10.7	9.5
					(6.6)	(6.2)*	(6.4)

Note: Standard errors are heteroscedasticity robust. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

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 specific students.

A final possible confounding factor is that if the rewards induce students in the reward groups to work harder, then this behavior could spill over to their peers in the control group. We consider it unlikely that spillover effects influence our findings. The overall pass rate of the students in our experiment—and, in particular, of the control group—is very similar to the pass rates of previous cohorts, and information about student effort from previous cohorts is in line with student effort among those who participated in our experiment. There is also no change in the composition of the student population in terms of secondary school grades for mathematics.

4. Effects on Achievement

The financial reward was tied to collecting all 60 credit points in the first-year. We therefore start by reporting the impact of treatment on the first-year pass rate and on the number of credit points students actually collected in the first year.

Table 4 shows the sample means for all students together and also by treatment and ability group. The first-year pass rate increases with the size of the reward from 0.195 in the control group to 0.202 and 0.241 in the small and large reward

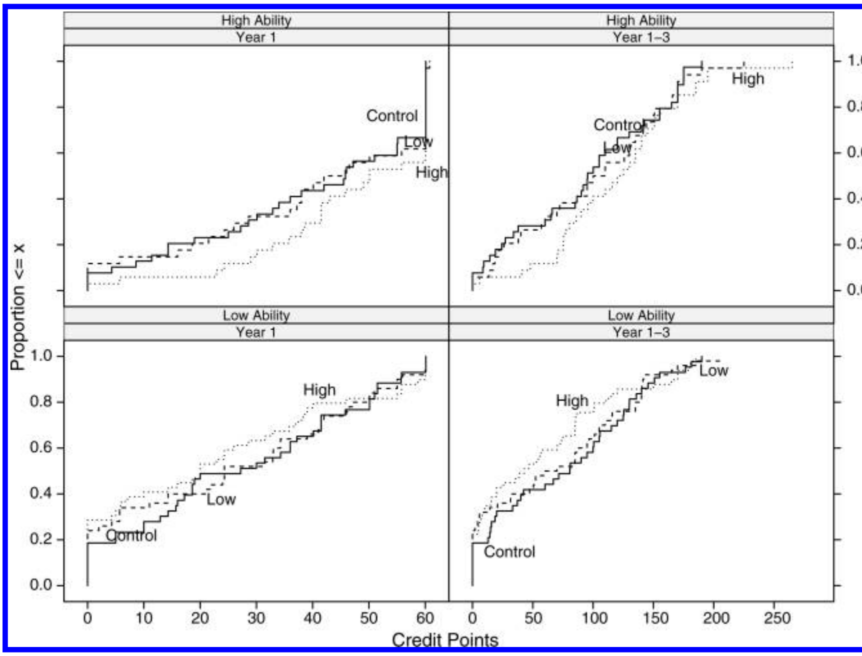

FIGURE 1. Cumulative distributions of credit points by treatment status for low- and high-ability students.

groups, respectively. This increasing pattern is present in both ability groups, although differences are not significant. Notice also the large difference in pass rates between high- and low-ability 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 (out of 60) credit points. The differences between the control and reward groups are small and not significant. Yet this finding, based on averages for the full sample, hides opposing effects for the high- and low-ability groups. The number of credit points collected in the high-ability group is increasing with the reward, whereas the converse holds in the low-ability group.

Figure 1 shows separately for the high-ability and the low-ability students the cumulative distribution functions of the number of credit points achieved after one year for the three treatment groups. The upper left-hand panel shows the first-year outcome distributions for high-ability students and the lower left-hand panel for low-ability students. It is immediately clear 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 outcome distribution for the large reward group is stochastically dominated by the outcome distribution for the small reward and control groups. Thus it seems the large reward has a positive incentive effect at the top of the ability distribution but 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.

The raw means in Table 4 and the graphs in Figure 1 both suggest strong interaction effects between reward size and ability. We now test this more formally by estimating linear regressions of the form

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

where D_i^S and D_i^L indicate that student i is assigned to the small and large reward group, respectively. For the descriptive results and the graphs in Figure 1, it was convenient to distinguish only two ability groups, but for the regression analyses we exploit all the ability variation available in the data. The variable A_i is thus the ability index based on the high school math score of student i, measured on a scale from 1 to 10. As additional regressors we include parents' education and the student's self-assessed pass probability (in the absence of a bonus). We include these controls which correct for remaining differences between groups and 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 10-point scale). The effect of a small reward for a student who scored a 10 (the highest possible score) on her high school math exam is then $\delta_S + 10\delta_S^A$; the effect for a (low-ability) student who scored a 4 is $\delta_S + 4\delta_S^A$.

Table 5 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 is 0.049 for the large reward. The pass rate thus increases with reward size, but the estimates are not significant. Similarly, there is no significant effect of the reward on the number of collected credit points.

Next we consider the interaction effects of reward size and ability in columns (2) and (4). These results confirm the descriptive results in Table 4 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 is larger for the large reward than for the small reward.

TABLE 5. First-year achievement with continuous ability interactions, OLS.

	Pass rate		Credit points	
	(1)	(2)	(3)	(4)
Small reward	0.014	−0.297	−0.929	−17.719
	(0.056)	(0.188)	(3.029)	(10.978)
Small reward × Ability		0.057		3.072
		(0.039)		(1.943)
Large reward	0.049	−0.327	−0.065	−32.798
	(0.059)	(0.203)	(2.941)	(10.155)***
Large reward × Ability		0.069		6.033
		(0.042)*		(1.764)***
Ability	0.120	0.082	6.704	4.006
	(0.017)***	(0.027)***	(0.883)***	(1.343)***
Schooling father	0.002	0.003	0.101	0.152
	(0.009)	(0.009)	(0.491)	(0.485)
Schooling mother	−0.005	−0.008	−1.049	−1.189
	(0.009)	(0.009)	(0.559)*	(0.563)**
Self-assessed pass prob.	0.247	0.266	21.334	22.440
	(0.085)***	(0.084)***	(4.939)***	(4.767)***
Intercept	−0.559	−0.345	−3.801	11.409
	(0.123)***	(0.162)**	(6.705)	(8.904)
F-test treatment (p-value)	*0.37 (0.689)*	*1.00 (0.410)*	*0.06 (0.944)*	*3.03 (0.018)*

Note: Regression estimates with heteroscedasticity robust standard errors in parentheses. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

We thus find that there is significant heterogeneity in the behavioral response to financial incentives that 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 yields both an incentive effect and a loss of intrinsic motivation. Whether the incentive effect is sufficiently large to compensate for 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, then the incentive effect will be zero and so the potential loss of intrinsic motivation dominates.[5] For students at the higher end of the skill distribution the threshold is feasible, and if the gap is positive then a positive incentive effect may dominate the loss of intrinsic motivation. This latter result is consistent with results reported in Angrist and Lavy (2009), who find that only those subjects in the upper part of the ability distribution respond to the rewards offered in their experiment.

5. This interpretation assumes that intrinsic motivation is reduced only by actual exposure to financial rewards and not by the mere announcement of the randomized experiment. If the announcement does have an effect then we assume that it is less than the effect of actual exposure. In that case, our estimates can be seen as a lower bound on the effect due to a loss of intrinsic motivation for low-ability students.

5. Effects on Effort and Time Allocation

The effect of rewards on achievement is a reduced-form effect. It disentangles neither the effects of rewards on effort nor subsequently those of effort on achievement. Furthermore, promising rewards might change study behavior. It may be, for example, that (low-ability) students in the control group focus on only a limited number of exams whereas similar students in the reward groups must succeed on all exams to have a chance of earning the reward. This might result in students in the reward groups having a lower probability of passing a particular exam than students in the control group. We therefore 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 for exams.)"
- "How many hours did you spend in total on preparing for re-take exams held in August? (Here we want to know the total number of hours, not the average per week.)"

Table 6 reports descriptive statistics on the various effort and time allocation measures by reward and ability groups. It shows that there are no substantial differences in exam-taking between the control group and the reward groups. High-ability students take about 11 to 12 exams including 2 re-take exams, whereas low-ability students take about 10 exams including 2.5 re-take exams. Among the low-ability students, those in the control group take the most exams. This rules out one 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, resulting in a high failure rate. The finding that low-ability students take fewer exams when assigned to a reward group is another indication that their intrinsic motivation is reduced by rewards. For high-ability students, those in the large reward group are slightly more likely to 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 preparing to re-take exams during the summer. Quite a few students report that they do not spend time at all on their study, which affects the averages for the second and third trimesters and for the summer period. These

TABLE 6. Descriptive statistics for time allocation by reward and ability group.

		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	10.85	11.20	10.35	11.02	11.56	11.15	12.71	10.86	9.80	9.86
	(5.56)	(5.37)	(5.57)	(5.75)	(4.11)	(4.51)	(3.42)	(6.33)	(6.18)	(6.71)
Re-take exams taken, 1st year	2.04	2.09	1.93	2.12	1.56	1.62	1.91	2.56	2.14	2.27
	(2.17)	(2.22)	(1.93)	(2.36)	(1.59)	(1.92)	(1.91)	(2.59)	(1.93)	(2.63)
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, with standard errors in parentheses.

are the students who dropped out and, for the summer period, those who did not re-take exams.[6] Students in the treatment (reward) groups tend to spend slightly less time on their studies, but the average time spent is similar across groups: differences neither substantial nor significant.

For high-ability students there are hardly any differences in study effort between those in the control group and both reward groups. For low-ability students, those in the high reward group devote on average less time to studying than students in the control group, which is consistent with the negative effects we find for low-ability students.

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

About 80% 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 €7.60 per hour. Here we see no differences between the reward and control groups, except that high-ability students in the large reward group work less than all other groups. Finally, the rewards did not deter students from joining a fraternity or from moving out of their parents' house.

We estimated 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, so total annual study effort is 14 times the weekly study effort in each of three terms plus total study effort during the summer. Because the re-take exams are given during two weeks in the summer, the total academic year consists of 44 weeks. Therefore, the average weekly study effort is the total annual study effort divided by 44. We also use the total number of exams taken, which is the sum of the regular exams and the re-take exams.[7]

Table 7 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. That we do not find significant effects of the rewards on students'

6. In the first trimester three respondents reported zero study effort; in the second trimester 33 respondents did so and in the third trimester 39. Zero hours were spent preparing for the August re-take exams by 83 students, of which 22 were not required to re-take exams. For the sample reporting positive numbers, the distribution of study time is bell shaped.

7. As indicated in footnote 6, 22 students passed all the regular exams and thus did not re-take any.

TABLE 7. Time allocation.

	Total effort		Exams taken	
	(1)	(2)	(3)	(4)
Small reward	−1.07	−5.03	−0.81	−6.56
	(1.55)	(6.57)	(0.84)	(3.66)*
Small reward × Ability		0.73		1.06
		(1.13)		(0.61)*
Large reward	−0.98	−5.19	−0.13	−7.87
	(1.61)	(6.52)	(0.83)	(3.63)**
Large reward × Ability		0.77		1.43
		(1.13)		(0.59)**
Ability	0.30	−0.18	0.58	−0.17
	(0.54)	(0.77)	(0.29)**	(0.44)
Schooling father	0.15	0.16	0.12	0.14
	(0.28)	(0.28)	(0.13)	(0.13)
Schooling mother	−0.59	−0.61	−0.41	−0.45
	(0.28)**	(0.29)**	(0.16)**	(0.16)***
Self-assessed pass prob.	8.16	8.47	3.58	3.93
	(2.84)***	(2.84)***	(1.49)**	(1.47)***
Intercept	18.54	21.17	−9.45	13.64
	(3.35)***	(4.47)***	(2.03)***	(2.81)***
F-test treatment (p-value)	*0.30 (0.739)*	*0.27 (0.897)*	*0.52 (0.596)*	*1.97 (0.100)*

Note: Regression estimates with heteroscedasticity robust standard errors in parentheses. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

self-reported effort is likely to be at least partly due to measurement error in the effort variables, which is also what Kremer, Miguel, and Thornton (2004) report. They find significantly positive effects of their rewards on observed school attendance but insignificant effects on self-reported measures of effort (and attitudes toward education).

Students in the reward groups take slightly fewer exams on average than students in the control group, but the differences are not significant. However, if we consider again the ability interaction, then the effects show the same pattern reported previously. High-ability students take more exams if assigned to a reward group, whereas low-ability students take fewer exams if assigned to one of the reward groups. These interaction effects are significant at the 10% level for the small reward and at the 5% level for the large reward. This, too, is consistent with rewards stimulating the study behavior of high-ability students while reducing the intrinsic motivation of low-ability students.

6. Performance Externalities

The rewards were tied to first-year performance only. It is nevertheless interesting to examine the impact of the rewards on students' achievement in the longer run. Opposing effects may be at work. On the one hand, students who worked harder when the incentive was in place may slow down afterward. On the other hand, good (respectively, bad) performance during the first-year may have a stimulating (respectively, discouraging) impact on effort and achievement in subsequent periods. Moreover, being on schedule after the first year may make subsequent studying easier because one need not re-take exams and is better prepared for more advanced courses. This section presents the effects of first-year rewards on achievement in the second and third year and on cumulative achievement after three years.

Table 8 shows the raw sample means of the number of credit points collected in the second and third year for all students together and by treatment and ability group. High-ability students collect about 30 credit points each year, whereas low-ability students earn about 20. For high-ability students the number of credit points collected each year increases with the size of the financial reward, but the pattern is decreasing for low-ability students. This pattern is identical to the one observed in the first year.

Table 8 also reports the cumulated number of credit points after three years and the dropout rate after three years. Again, the same picture emerges: High-ability (respectively, low-ability) students collected more (respectively, fewer) credit points when exposed to a larger financial incentive. Results for the dropout rate reiterate this pattern: High-ability students are less likely to drop out when they could earn a larger reward in the first-year, and low-ability students are more likely to have dropped out if assigned to one of the treatment groups. The drop-out rate is lowest for high-ability students in the large reward group and is highest for low-ability students in the large reward group.

We continue by presenting results of estimating equation (1), now using credit points in the second year, credit points in the third year, cumulated credit points after three years, and the dropout rate after three years as the dependent variables. The odd-numbered columns in Table 9 report results from specifications without interaction effects of reward size and ability (measured on a 10-point scale); the even numbered columns are based on specifications that include these interactions. Although the results in the columns without interaction effects suggest that financial rewards have no significant impact on later outcomes, the results in the columns that include interaction effects reveal that this zero average impact hides heterogeneous treatment effects. The interaction terms affect achievement positively, and this positive effect increases with the size of the reward.

The pattern of performance spillovers reported in this section indicates that high-ability students who were induced by the financial incentive to perform better in the first year do not slow down in subsequent years; instead they benefit from the positive feedback received in the first year and from not having to re-take (many) exams. In contrast, low-ability students whose first-year work was adversely affected by the impact of the reward scheme on their intrinsic motivation seem to have become discouraged by their poor performance and the resulting lack of positive feedback.

TABLE 8. Achievement externalities (2nd and 3rd year) and credit points and drop-out (after 3rd year).

		Treatment								
	Mean	Control (1)	Low (2)	High (3)	(2) − (1) (4)		(3) − (1) (5)		(3) − (2) (6)	
A. Credit points, 2nd year										
- All	25.0	26.1	23.9	25.0	−2.2	(3.6)	−1.1	(3.7)	1.1	(3.7)
- High-ability	31.1	28.4	28.4	36.8	0.0	(5.4)	8.4	(5.5)	8.4	(5.7)
- Low-ability	20.4	24.0	20.8	16.7	−3.2	(4.8)	−7.3	(4.8)	−4.1	(4.4)
DID (high- vs. low-ability)					3.3	(7.2)	15.7	(7.3)**	12.4	(7.2)*
B. Credit points, 3rd year										
- All	23.1	22.3	24.3	22.7	2.0	(3.5)	0.4	(3.6)	−1.6	(3.6)
- High-ability	27.6	23.2	29.0	31.2	5.8	(5.3)	8.0	(5.5)	2.2	(5.8)
- Low-ability	19.7	21.5	21.1	16.8	−0.4	(4.7)	−4.6	(4.7)	−4.3	(4.5)
DID (high- vs. low-ability)					6.2	(7.1)	12.7	(7.2)*	6.4	(7.3)
C. Total credit points (year 1–3)										
- All	83.1	84.3	81.8	83.1	−2.5	(9.6)	−1.2	(9.8)	1.3	(9.9)
- High-ability	103.9	94.5	100.0	118.7	5.6	(14.3)	24.2	(13.7)*	18.7	(14.3)
- Low-ability	67.3	75.0	69.5	58.4	−5.6	(12.6)	−16.6	(12.5)	−11.1	(12.2)
DID (high- vs. low-ability)					11.1	(19.1)	40.9	(18.5)**	29.7	(18.8)
D. Drop-out rate										
- All	0.365	0.402	0.345	0.349	−0.057	(0.075)	−0.053	(0.076)	0.004	(0.074)
- High-ability	0.243	0.359	0.235	0.118	−0.124	(0.107)	−0.241	(0.096)**	−0.118	(0.093)
- Low-ability	0.458	0.442	0.420	0.510	−0.022	(0.104)	0.068	(0.105)	0.090	(0.101)
DID (high- vs. low-ability)					−0.102	(0.149)	−0.310	(0.142)**	−0.208	(0.137)

Note: Standard errors are heteroscedasticity robust. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

TABLE 9. Later achievement with continuous ability interactions, OLS.

	Credit points, 2nd year		Credit points, 3rd year		Credit points, year 1–3		Drop-out	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Small reward	−2.369	−20.996	1.859	−23.581	−2.189	−64.618	−0.066	0.218
	(3.453)	(13.642)	(3.413)	(13.921)*	(8.962)	(34.505)*	(0.074)	(0.284)
Small reward × Ability		3.405		4.670		11.434		−0.052
		(2.498)		(2.497)*		(6.244)*		(0.049)
Large reward	−0.760	−41.426	0.933	−29.944	0.087	−108.259	−0.056	0.668
	(3.598)	(13.014)***	(3.514)	(14.462)**	(8.959)	(32.814)***	(0.073)	(0.277)**
Large reward × Ability		7.497		5.681		19.960		−0.134
		(2.358)***		(2.639)**		(5.874)***		(0.046)***
Ability	4.551	1.341	4.240	1.090	16.089	6.728	−0.078	−0.024
	(1.134)***	(1.811)	(1.098)***	(1.747)	(2.785)***	(4.395)	(0.022)***	(0.036)
Schooling father	0.494	0.546	0.181	0.273	0.895	1.094	−0.002	−0.002
	(0.557)	(0.553)	(0.532)	(0.535)	(1.441)	(1.437)	(0.012)	(0.012)
Schooling mother	−1.575	−1.736	−1.351	−1.542	−4.040	−4.547	0.014	0.017
	(0.620)**	(0.607)***	(0.620)**	(0.628)**	(1.627)**	(1.620)***	(0.013)	(0.013)
Self-assessed pass prob.	10.285	11.545	10.512	12.047	43.538	47.551	−0.323	−0.343
	(5.868)*	(5.625)**	(5.813)*	(5.617)**	(15.157)***	(14.444)***	(0.124)***	(0.121)***
Intercept	8.246	26.403	7.432	24.958	9.803	62.415	0.855	0.548
	(8.209)	(11.406)**	(8.975)	(11.926)**	(21.860)	(28.758)**	(0.164)***	(0.229)**
F-test treatment (p-value)	*0.25 (0.780)*	*2.67 (0.033)*	*0.15 (0.862)*	*1.47 (0.213)*	*0.04 (0.959)*	*2.92 (0.022)*	*0.46 (0.635)*	*2.81 (0.026)*

Note: Regression estimates with heteroscedasticity robust standard errors in parentheses. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

7. Conclusion

This paper reports on 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 it was €227. Students in the control group did not earn a reward.

We find that the average effects of the rewards on the first-year pass rate are small and not statistically significant. There are no average effects on the number of achieved credit points by the end of the first year. Further breakdown of these results suggests that there is significant heterogeneity in the behavioral response to the financial incentives. High-ability students have higher pass rates and collect significantly more credit points when assigned to larger reward groups. In contrast, low-ability students appear to achieve less when assigned to larger reward groups. After the first-year these effects are significant only for the high-ability group, but after three years effect sizes have increased and are statistically significant for both the low- and high-ability group.

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

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

References

Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer (2002). "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment." *American Economic Review*, 92(5), 1535–1558.

Angrist, Joshua D., Daniel Lang, and Philip Oreopoulos (2009). "Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement." *American Economic Journal: Applied Economics*, 1(1), 136–163.

Angrist, Joshua D., and Victor Lavy (2002). "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." NBER Working Paper No. 9389.

Angrist, Joshua D., and Victor Lavy (2009). "The Effect of High-Stakes High School Achievement Awards: Evidence from a School-Centered Randomized Trial." *American Economic Review*, 99, 1384–1414.

Camerer, Colin F., and Robin M. Hogarth (1999). "The Effects of Financial Incentives in Experiments: A Review and Capital-labor-production Framework." *Journal of Risk and Uncertainty*, 19, 7–42.

Dearden, Lorraine, Carl Emmerson, Christine Frayne, Alissa Goodman, Hidehiko Ichimura, and Costas Meghir (2001). "Education Maintenance Allowances: The First Year—A Quantitative Evaluation." Technical Report RR257, Department for Education and Skills, London.

Dearden, Lorraine, Carl Emmerson, Christine Frayne, and Costas Meghir (2002). "Education Maintenance Allowances: The First Two Years—A Quantitative Evaluation." Report RR352, Department for Education and Skills, London.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore (2007). "College Cost and Time to Obtain a Degree: Evidence from Tuition Discontinuities." CEPR Discussion Paper No. 6106.

Hilkhuysen, Peter (2000). "Premie Studieresultaat; Experiment in de Propedeuse AEO: Cohort 1999/2000." Report, Faculteit der Economische Wetenschappen en Econometrie, Universiteit van Amsterdam.

Kremer, Michael, Edward Miguel, and Rebecca Thornton (2004). "Incentives to Learn." NBER Working Paper No. 10971.

Leuven, E., H. Oosterbeek, and B. Van Der Klaauw (2003). The Effect of Financial Rewards on Students' Acheivements: Evidence from a Randomized Experiment." CEPR Discussion Paper No. 3921.