Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment

Author(s): Maria De Paola, Vincenzo Scoppa, and Rosanna Nisticó
Source: *Journal of Human Capital*, Vol. 6, No. 1 (Spring 2012), pp. 56-85
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/664795
Accessed: 18/12/2014 15:22

Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment

Maria De Paola
University of Calabria

Vincenzo Scoppa
University of Calabria

Rosanna Nisticò
University of Calabria

We evaluate the effectiveness of monetary incentives in enhancing student performance using a randomized experiment involving undergraduate students enrolled at a southern Italian University. Students were assigned to three different groups: a high-reward group, a low-reward group, and a control group. Rewards were given to the 30 best-performing students in each group. Financial rewards increase student performance. High-ability students react strongly whereas the effect is null for low-ability students. Large and small rewards produce very similar effects. These effects also persist in subsequent years, when the financial incentives are no longer in place. No types of crowding-out effects of the monetary incentives are found.

I. Introduction

The use of monetary incentives to improve student performance is becoming a common practice in the educational systems of several countries. An increasing number of programs are being implemented in developing countries. For example, in Colombia, the Familias en Acción

The project was funded by the regional government of Calabria (Assessorato Istruzione, Alta Formazione e Ricerca, Piano Operativo Regionale Calabria, European Social Fund 2007–13). We would like to thank the University of Calabria for providing access to the administrative data and Alfio Cariola, Domenico Cersosimo, Alfredo Fortunato, and Sara Laurita, who helped with the project. We have also benefited from helpful comments from the editors Eric Hanushek and Isaac Ehrlich, two anonymous referees, Mariarosaria Agostino, Giovanni Anania, Lorenzo Cappellari, Federica Demaria, and Michela Ponzo. We also thank seminar participants at the First International Workshop on Applied Economics of Education, Catanzaro, 2010, and participants in the seminars held at the University of Cagliari and at the University of Calabria.

[*Journal of Human Capital,* 2012, vol. 6, no. 1]
 1932-8575/2012/0601-0003$10.00

and Programa de Ampliación de Cobertura de Educación Secundaria programs assigned vouchers, respectively, to primary and secondary school students, in relation to attendance and performance (Angrist et al. 2002); in Mexico, the PROGRESA financial incentive scheme aimed at increasing school enrollment and attendance offered cash payments to families (Schultz 2004); in Kenya, the Girls' Scholarship Programme provided awards to female students who obtained high test scores (Kremer, Miguel, and Thornton 2009).

Similar programs have also been recently implemented in a number of developed countries. In the United States, following the Helping Outstanding Pupils Educationally program implemented in Georgia, many others states have introduced analogous scholarship programs (Heller 2006). About 27,000 students benefited from the financial incentives offered by a number of schools in Chicago, Dallas, and New York (Fryer 2011).

A conspicuous body of literature has evaluated financial incentive schemes. Results suggest that incentives work well in increasing attendance and enrollment, especially in developing countries (Angrist et al. 2002; Angrist, Bettinger, and Kremer 2006; Kremer et al. 2009), whereas the findings are ambiguous with regard to the impact on students' effort and achievement (Angrist, Lang, and Oreopoulos 2009; Bettinger 2010; Levitt, List, and Sadoff 2010; Rodriguez-Planas 2010; Fryer 2011). Furthermore, a wide heterogeneity of effects emerges from these programs according to students' characteristics (Angrist and Lavy 2009; Croson and Gneezy 2009; Bettinger 2010).

Most of these experiments involved primary and secondary school students; only the experiments implemented by Angrist et al. (2009) and Leuven, Oosterbeek, and van der Klaauw (2010) were oriented toward college students.

Angrist et al. (2009) conducted an experimental evaluation of different strategies aimed at improving student academic performance at a large Canadian university. Students were randomly assigned to three treatment groups: a number of support services were offered to the first group, financial rewards for good performance were offered to the second, and a combination of support services and monetary incentives was offered to the third. The results of the experiment show that females improved their academic performance, whereas males did not react to any of the proposed strategies. The positive effect of the combination of support services and monetary incentives also continued into the following year when the program was no longer available.

In an experiment conducted by Leuven et al. (2010), students at the University of Amsterdam were randomly assigned to two treatment groups: a high-reward group to which a bonus of €681 was promised for completing all first-year exams and a low-reward group to which a bonus of €227 was promised. The empirical analysis shows a positive effect on high-ability students assigned to the large-reward group but a negative effect on low-ability students. These effects tend to persist over time even when the rewards are removed.

The findings emerging from these works, showing small and heterogeneous effects, do not lead to clear implications regarding the role of financial incentives in increasing the skills acquired by undergraduate students. One factor that may reduce the effectiveness of financial incentives is the negative impact that they may produce on intrinsic motivation (Gneezy, Meier, and Rey-Biel 2011). Related to the undermining effect extrinsic rewards may have on intrinsic motivation is the concern that individuals who receive concrete rewards for doing something they already did in the absence of such incentives will do less after the incentives are removed. However, as suggested by the persistent effects emerging from Angrist et al. (2009) and Leuven et al. (2010), it could also be the case that extrinsic rewards help individuals establish positive patterns that will be maintained after the incentives are removed. Clearly, these positive patterns will make monetary incentives more appealing than they would be if the positive behavior lasted only as long as the incentives did. In fact, a crucial issue in the evaluation of monetary incentive programs is their cost-effectiveness. Financial incentives are typically quite expensive, and positive effects on achievement would have to justify these costs in comparison with other feasible alternatives.

In this paper, we investigate the effects of monetary incentives on student performance in a southern Italian region, an area where—because of the imperfections of the labor market—individuals may lack incentives to provide effort in order to acquire skills. The low rewards offered to skills may discourage students from studying hard (De Paola and Scoppa 2007) and may, at least in part, be responsible for the high dropout rates and the excessive duration of students' academic careers. The impact of financial incentives may be stronger in such an environment since, as we show in the Appendix to the paper, the corresponding increase in the marginal return to effort is much more pronounced in labor markets that pay low returns to skills.

The experiment involved 462 first-year undergraduate students enrolled in the business administration degree course offered by the University of Calabria in the academic year 2008–9. Students were assigned, through a stratified random procedure, to one control group and to two "treatment groups," eligible for the attainment of a small (€250) or a large (€700) monetary prize on the basis of their academic performance. The latter feature of the experiment was aimed at disentangling the effect that derives from the simple existence of an incentive from that related to its size. Prizes were assigned to the 30 best-performing students in each treated group (corresponding to the top 20 percent students) in terms of credits earned and grades obtained on examinations.

Although policy interventions in education are typically aimed at helping disadvantaged students, for a number of reasons, we chose to give prizes just to the best students. First, paying a reward only to the top 20 percent students entails low costs for the public budget, and evaluating the effects of such a policy was of particular interest in a region with few resources available for public policies. Second, even though the program was more likely to affect students endowed with high abilities than low-ability students, from a policy point of view, it is questionable whether helping low-ability students to graduate has higher social returns than improving the skills of high-ability students. Increasing the skills of more promising students, especially in depressed areas, may lead to high social returns as these students may play an important role in creating a new ruling class. Third, students entering college might not be fully aware of their academic abilities, could be overly optimistic about their ability (Stinebrickner and Stinebrickner 2009; Arcidiacono, Hotz, and Kang 2012), and so might all be induced to provide effort in an attempt to achieve the experiment target.

The design of our experiment is purposely similar to the experiment undertaken by Leuven et al. (2010) with the aim of evaluating whether, because of cultural and economic differences between geographical areas, similar incentive schemes are perceived differently and produce different outcomes.

From our analysis it emerges that financial rewards are effective in enhancing student performance. The prospective of obtaining an award of €700 significantly increases student performance (by about 0.19 standard deviations of the target variable). Competing for an award of €250 produces an effect that is slightly smaller (about 0.16 standard deviations). Treated students improve their performance in terms of both credits earned and grades obtained (as required by the experimental design), but not at the expense of a poorer performance in examinations not included in the experiment.

The effects are heterogeneous between high- and low-ability students: high-ability students in the treated groups perform significantly better than students of comparable ability in the control group, whereas low-ability students in the treated and control groups do not show any significant difference in performance.

We find that high-ability students also benefit from the program in subsequent years when the financial incentives are no longer in place. The effect of the program on low-ability treated students continues to be null: their performance in subsequent years is indistinguishable from the performance of similar students in the control group. Therefore, it seems that any eventual loss of intrinsic motivation relating to the introduction of financial incentives did not prevail either for high-ability students or for low-ability students who were unlikely to achieve the experiment target. We find that large and small rewards produce very similar effects, both in the short and in the long run, implying that the size of the prize is not particularly relevant for enhancing student effort.

The paper is organized as follows. In Section II, the design of the experiment is explained and some information about the Italian university system is provided. Section III discusses the effects of the financial rewards on credits and grades earned by students. In Section IV, we analyze the treatment effects using student-course-level observations. In Section V, using an instrumental variable strategy, we investigate the treatment effects on students who have effectively participated in the experiment. Section VI discusses the answers given by students in a postexperiment survey. Section VII examines students' pass rates in non-target examinations. Section VIII is devoted to analyzing the impact of the experiment on student performance in subsequent years when the financial incentives were no longer in place. Section IX presents concluding remarks.

II. Background, Experiment Description, and Data

Italian labor markets are far from being competitive, and as has been widely discussed in the literature, distortions are even larger in the South of the country. In the South, effective skills are not adequately rewarded because of the importance that family background and social networks play in local labor markets (Fabbri and Rossi 1997); the relevance of public-sector employment, where meritocratic approaches to pay and promotions are rarely applied (see Alesina, Danninger, and Rostagno 2001); and the industrial system, which specializes in traditional and technologically poor sectors requiring low levels of skill. Distorted labor markets and low student effort may also have a role in explaining the huge differences between the results of students from the North and the South of Italy in international tests, such as the OECD's Programme for International Student Assessment or Trends in International Mathematics and Science Study.

The regional government of Calabria, aware of the low level of basic competences acquired by students living in the area and of the implications this has on social and economic development, has planned to introduce a program awarding scholarships to students with high achievements, with the financial support of the European Union (through the European Social Fund). The randomized experiment presented in this paper was funded as part of this policy intervention so as to evaluate whether financial incentives and merit scholarships might be effective tools in improving student performance.

As we will show in the Appendix, where we present a very simple model of student effort and incentives, financial rewards are arguably more effective in enhancing student performance in environments where there are low returns to skills.

The experiment we conducted involved 462 students enrolled in the first year of the first-level degree course in business administration (Laurea Triennale in Economia Aziendale) at the University of Calabria in the academic year 2008–9.[1] The University of Calabria is a middle-sized public university located in the South of Italy. It has currently about 34,000 students enrolled in different degree courses and at different levels of the Italian university system, which, since the reform of 2001, has been organized into three main levels: first-level degrees (legal duration of 3 years), second-level degrees (2 further years), and PhD degree. In order to gain a first-level degree, students have to acquire a total of 180 credits. Students who have acquired a first-level degree can undertake a second-level degree.[2]

The first-level degree course in business administration is highly in demand and takes on a large number of students.[3] The first year includes six compulsory courses: (1) calculus, (2) accounting, (3) statistics, (4) microeconomics (10 credits each), (5) public law, and (6) business administration (five credits each). In addition, students have to pass three additional examinations (English, French, and computer sciences), worth a total of 10 credits, which we do not include in our target variable (see below). In the first year, all students have to take the same classes and cannot choose any of their courses.[4]

Given the high number of students enrolled, they are assigned, on the basis of the alphabetic order of their surnames, to two different teaching classes (class 1 [A–L] and class 2 [M–Z]) for all their compulsory first-year courses. Students in the two teaching classes attend courses in the same term with the same programs.

At the beginning of the academic year, all the students were informed of the experiment through both presentations during classes and a letter explaining the purpose and the format of the experiment.[5]

On the basis of the available administrative information on students' characteristics, we proceeded with the stratification of students according to the following variables: gender, class in which students attend courses (classes 1 and 2), type of high school attended (three categories: lyceum, technical schools, vocational and other types of schools), and grade obtained at the final high school examination (split into four categories corresponding to quartiles). Following this procedure, we obtained 47 (out of 48) nonnull groups. Within each group, one-third of the students were randomly assigned to the large-reward group ("treatment A" with a reward of €700), one-third to the low-reward group

[1] Budget constraints did not allow us to extend the experiment to students enrolled in other degree courses.

[2] After having accomplished their second-level degree, students can enroll in a PhD course.

[3] Applications exceeded the number of students to be enrolled, although it is the degree course with the largest number of enrolled students at the University of Calabria.

[4] In each academic year, Italian students have several dates available to take a given exam (in fact, at the University of Calabria, three sessions are available). Retake exams are possible in case of failure.

[5] Moreover, a Web page on the university Web site provided students with information regarding the experiment.

("treatment B" with a reward of €250), and one-third to the control group.[6] We ended up with 155 students assigned to treatment A, 150 to treatment B, and 157 to the control group.[7]

To give an idea of the real size of the rewards, consider that the average monthly wage a student could earn on the labor market is about €900 (€5.6 per hour),[8] meaning that the reward of €700 is equivalent to 125 hours' work; the reward of €250 is the equivalent of about 45 hours' work. Given the very low wages students could obtain in the labor market and also considering that the probability of finding a job is quite low in the area in which they live, it is likely that students considered the work needed to earn the reward adequate to compensate them for forgone wages.

In order to have an indicator of student awareness and interest in the experiment, we asked students to sign a list indicating their treatment status, and we announced that students who did not comply were ineligible for the prizes. Approximately 91 percent of students in treatment A and 88 percent of students in treatment B filled in their participation form.

Prizes were promised to the 30 best-performing students in each treatment group. Given the low percentage of students rewarded, the cost of the program was quite small. Therefore, one advantage of the policy we study is that it can also be implemented in situations in which public budget constraints do not allow many resources to be devoted to educational policies.

One potential disadvantage of this type of policy lies in the fact that the program is more likely to affect students endowed with high abilities rather than low-ability students. Nevertheless, there were good ex ante reasons to believe that the program may also have been effective in increasing the abilities of students at the lower tail of the ability distributions, since our students tended to be overly optimistic with regard to their abilities. According to the answers given by sample students, at the moment of their enrollment, to a survey question asking them about their expectations regarding the probability of their getting their degree within the time foreseen by the university,[9] about 94 percent of them (94.7 percent of the high-ability students and 93.6 percent of the low-ability students) estimated this probability to be higher than 50 percent.

[6] When the number of students included in each stratified group was not a multiple of three, one or two students were assigned randomly to the treatment or control groups.

[7] The random assignment procedure was carried out in the presence of students. Students were also informed by e-mail of their assignment status, and the list of students belonging to each group was published on the experiment Web page.

[8] To obtain information on the average wage that sample students could gain on the labor market, we used data from the Bank of Italy Survey of Household Income and Wealth considering just those individuals living in the South and in the 19–25 age range.

[9] All students applying for a place at the University of Calabria were invited to participate in an online survey asking a number of questions on individual characteristics, family background, motivation, expectations, etc. We use the answers to the question "What probability would you give to your obtaining your degree within the regular time?"

About 44 percent of the students (43.9 percent of the high-ability students and 44.9 percent of the low-ability students) stated that this probability was higher than 80 percent. These statements appear excessively optimistic since, 3 years later, only one of the interviewed students had actually concluded his or her degree program. On average, the students have acquired 124 of the 180 credits needed to complete their degree program. As a consequence, they are, on average, a year behind in their academic career.

To avoid the possibility of the incentive system generating dysfunctional responses, deriving from the attempt by students to maximize their achievements with regard to their target at the expense of other objectives ("multitasking" in the definition of Holmstrom and Milgrom [1991]), student performance was measured considering both grades obtained at examinations and the number of credits earned. To be more precise, our measure of targeted performance, *total points*, is equal to the sum of the grades obtained in each class.[10] In this way, we avoided both the risk that students aim to pass as many examinations as possible, albeit with low grades, and the risk that they might seek to obtain high grades by concentrating on only a few examinations.

Pass grades range from 18 to "30 cum laude," which we consider equal to 31. Only exams undertaken by July 31, 2009, were taken into account in determining student performance. Most of the examinations included in our target variable are based on multiple-choice tests. For these examinations (calculus, accounting, statistics, and microeconomics), we have also considered grades obtained in failed examinations (below the minimum pass line of 18/30) in defining our target variable. On the other hand, for examinations in which grades below the minimum pass line were not recorded (public law and business administration), we have considered a grade equal to zero as a fail.[11]

Table 1 provides descriptive statistics for the sample of students. About 55 percent of the students were female. The average age was 20.5. High school grade ranged from 60 (the minimum pass grade) to 100 (the maximum grade), with a mean of 89.4. Students mainly came from scientific lyceums (about 33 percent) and technical schools (about 40 percent); about 21 percent attended a vocational school and 6 percent a classical lyceum. About 57 percent of the students attended lectures in class 1 (surnames from A to L). At the end of July, students had earned 18 credits on average, with an average grade in passed examinations of 22.4. The target variable total points has a mean of 43.3 with a standard deviation of 32.

In columns 1–3 of table 2, means for a number of individual char-

[10] We have divided by two the grades obtained in classes to which five credits are assigned (business administration and public law).

[11] We presented the experiment to the instructors, explaining the aims of the project and asking them for complete impartiality in grading policy (they were not informed of students' treatment status).

TABLE 1
Descriptive Statistics

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Female	462	.554	.498	0	1
Age	462	20.515	2.514	19	51
High school grade	462	89.435	8.737	60	100
Scientific lyceum	462	.325	.469	0	1
Classical lyceum	462	.061	.239	0	1
Technical schools	462	.400	.491	0	1
Vocational schools	462	.214	.411	0	1
Class 1: A–L	462	.569	.496	0	1
Total points	462	43.286	32.376	0	139.5
Credits	462	18.468	12.768	0	49
Average grade	383	22.402	2.554	18	31

Note.—Grades for each course range from 18 to "30 cum laude" (set equal to 31). High school grade ranges from 60 to 100.

acteristics are reported, by treatment groups. Differences in means between treatment A and control and treatment B and control are presented in columns 4 and 5, respectively (standard errors are reported in parentheses). In column 6, we report the F-statistic (and p-value) for a test of equality of variable means across all three groups.

Results show that the randomization was successful in creating comparable treatment and control groups with regard to the observable characteristics: there are no significant differences between the treatment statuses in terms of students' sex, age, high school grade, type of high school attended, and classes in which they attended first-year courses.

III. Empirical Results

In this section, in order to analyze the effects of financial incentives on student performance, we estimate the following model by ordinary least squares (OLS):

$$Y_i = \beta_0 + \beta_1(\text{treatment } A_i) + \beta_2(\text{treatment } B_i) + \phi X_i + u_i,$$

where Y_i is a measure of performance (alternatively, total points or credits) of student i; X_i is a vector of individual characteristics (gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence,[12] an interaction term between high school grade and lyceum); treatment A_i is a dummy variable that takes a value of one if student i has been assigned to the group competing for an award of €700 (and zero otherwise); treatment B_i is a dummy variable that takes a value of one if student i has been assigned to the group competing for an award of €250 (and zero otherwise); and

[12] Provinces correspond to the Nuts 3 Eurostat classification.

TABLE 2

STUDENT CHARACTERISTICS ACROSS TREATMENT AND CONTROL GROUPS

	Means			Differences		
	Treatment A (1)	Treatment B (2)	Control (3)	Treatment A vs. Control (4)	Treatment B vs. Control (5)	*F*-Statistic (*p*-Value) (6)
Female	.529	.580	.554	−.025	.026	.399
				(.057)	(.057)	(.671)
Age	20.723	20.360	20.459	.264	−.099	.853
				(.320)	(.228)	(.427)
High school grade	89.432	89.060	89.796	−.364	−.736	.271
				(.982)	(.993)	(.762)
Scientific lyceum	.323	.300	.350	−.028	−.050	.443
				(.054)	(.054)	(.642)
Classical lyceum	.065	.080	.038	.026	.042	1.206
				(.025)	(.027)	(.300)
Technical schools	.400	.393	.408	−.008	−.014	.033
				(.056)	(.056)	(.968)
Vocational and other schools	.213	.227	.204	.009	.023	.119
				(.046)	(.047)	(.887)
Class 1: A–L	.568	.573	.567	.001	.006	.008
				(.056)	(.057)	(.992)

Note.—Standard errors are reported in parentheses. In col. 6, we report the *F*-statistic and *p*-value for a test of equality of variable means across all three groups.

TABLE 3
The Impact of the Incentives on Student Performance (Measured by Total Points): OLS Regressions

	All (1)	Top (2)	Bottom (3)	Males (4)	Females (5)
Treatment A (€700)	6.023**	10.783**	1.435	5.390	5.841
	(3.059)	(4.760)	(3.819)	(4.615)	(4.061)
Treatment B (€250)	5.350*	10.697**	−.088	2.354	6.157
	(3.164)	(4.557)	(4.146)	(4.877)	(4.207)
Female	−.531	−6.086	3.474		
	(2.644)	(4.074)	(3.163)		
Age	−1.217**	−5.215**	−1.072**	−.388	−.865
	(.516)	(2.615)	(.448)	(.754)	(1.108)
High school grade	1.224***	3.256***	.389*	1.105***	1.398***
	(.203)	(.703)	(.227)	(.297)	(.262)
Scientific lyceum	31.815***	29.188***	7.368	39.358***	26.287***
	(3.550)	(5.833)	(8.893)	(5.419)	(4.890)
Classical lyceum	13.424**	6.707	6.858	−1.556	13.422*
	(6.522)	(9.587)	(11.065)	(13.145)	(7.764)
Technical schools	15.429***	.000	5.966	22.129***	9.487*
	(3.398)	(.000)	(3.673)	(4.835)	(4.940)
Class 1: A–L	1.630	4.753	−2.224	−1.370	3.968
	(2.585)	(3.806)	(3.503)	(4.171)	(3.405)
Lyceum × high school grade	.845***	−.785	−.120	.965**	.773*
	(.290)	(.778)	(.612)	(.487)	(.400)
Constant	−65.112***	−163.519**	9.441	−75.677**	−84.283**
	(20.464)	(77.225)	(21.903)	(30.190)	(36.264)
Observations	462	231	231	206	256
R^2	.327	.252	.136	.380	.317

Note.—The dependent variable is total points. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions, we control for province of residence dummies (six categories, not reported).
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

u_i is an error term capturing idiosyncratic shocks or unobserved student characteristics.

The coefficients β_1 and β_2 capture the effects of the financial incentives: β_1 measures the average impact on student performance of competing for an award of €700 (with respect to the control group), and the coefficient β_2 measures the average impact of competing for an award of €250.

We first focus on total points, the target variable on the basis of which the merit scholarships were assigned. OLS estimates are reported in table 3. In all the specifications, standard errors are robust to heteroskedasticity.

Our main specification is reported in column 1. Ceteris paribus, the prospective of obtaining an award of €700 increases student performance by 6.02 points, that is, 0.19 standard deviations of total points. The coefficient is significant at the 5 percent level. Competing for an award of €250 increases performance by 5.35 points, 0.16 standard deviations of total points, significant at the 10 percent level and slightly below the effect produced by treatment A.

The effects of control variables are consistent with the findings presented in the existing literature. Students with a higher high school grade show a much better academic performance. In addition, students who attended a lyceum perform better than students who come from technical or vocational schools. We have also included an interaction term between high school grade and lyceum among our regressors, which turns out to be positive and highly significant, implying that the high school grade increases academic performance significantly more for lyceum students. Older students perform significantly worse. The dummy female is not always statistically significant, and there are no significant differences between the two different first-year teaching classes that students attended.

In order to evaluate the magnitude of the estimated treatment effects, it is useful to consider that the effect of being included in treatment A corresponds to the effect produced by an increase in student ability, measured by high school grade, of about 3 points for students who attended a lyceum and of 5 points for students who attended a technical or vocational school.

A typical criticism of the type of experiment we have implemented is that individuals who are not selected for the reward groups (i.e., those in the control group) might somehow feel cheated and decrease their effort relative to a case in which no incentive scheme was in place. In this case, our estimates would overstate the effects of the program on the treated groups. On the other hand, our estimates will suffer a downward bias if students in the control group have increased their effort because they are afraid of losing ground to students in the other groups who have an incentive to work hard.

We have collected data on the performance of students enrolled in the first-level degree course in business administration in the academic year 2009–10 (the year after our experiment)[13] with the aim of evaluating the relevance of these problems for our results. Focusing on all the 2009–10 students and on students in the control group enrolled in 2008–9, we regress our measures of performance (total points and credits) on a dummy for students enrolled in 2008–9 (results are not reported). We do not find any statistically significant difference between the performance of control students and students enrolled in 2009–10. This result holds true also when, to control for differences between the two cohorts, we include among our regressors a number of student

[13] We could not use the previous cohort since, because of an institutional change that took place in 2008, the 2007–8 and the 2008–9 cohorts are not comparable. In 2008, some new university rules were introduced in Italy that imposed a limit on the maximum number of exams needed to gain a first-level degree and organized teaching activity (previously organized into short modules, 30–40 hours of teaching activity) into long modules with about 60–80 hours of teaching activity.

characteristics such as gender, high school grade, and type of high school attended. This reassures us that students included in the control group did not react to the experiment. However, we are not able to exclude the existence of unobservable differences among cohorts, which may be canceled by the reactions of students assigned to the control group.

Given the structure of the experiment, with a competition among students in which only the best-performing ones in each treated group were rewarded, we investigate whether high-ability students reacted differently from low-ability students. With this aim, we follow the approach of Angrist and Lavy (2009) and use the predicted performance of students (in terms of both total points and credits) in order to obtain a composite measure of student ability and to divide students into two equal-sized groups. We first estimate a model for student performance by considering students in the control group exclusively, using as explanatory variables the vector of individual characteristics X_i, including gender, age, type of high school attended, and high school grade. The estimated coefficients of this model, denoted with the vector $\hat{\phi}$, are then used for predicting the performance of students in treated and control groups, on the basis of their effective characteristics:

$$\hat{Y}_i = \hat{\phi}_0 + \hat{\phi}_1 X_{1i} + \cdots + \hat{\phi}_k X_{ki}.$$

We consider students with $\hat{Y}_i \geq \text{median}(\hat{Y})$ to be of high ability and those with $\hat{Y}_i < \text{median}(\hat{Y})$ to be low-ability students.

The effects on students with a predicted ability above or below the median are reported, respectively, in columns 2 and 3 of table 3. It emerges that the effect of financial rewards is negligible for lower-ability students, whereas both the small and the large rewards produce a positive effect for higher-ability students. The effects of the two rewards are very similar, quite large (about 10.7 points more, an increase of 0.33 standard deviations), and significant at the 5 percent level.

Therefore, high-ability students in the treated groups perform significantly better than students of comparable ability in the control group. On the other hand, low-ability students in the treated and control groups do not show any significant difference in performance. The null effect on low-ability students suggests that even though, at the moment of their enrollment, they had very optimistic expectations about their performance (their expectations were very similar to those of high-ability students, as was discussed in Sec. II), they change their opinions during the first year of the degree program and become more realistic about their relative abilities. As a consequence, when facing a particularly demanding target, they appear discouraged from providing the significant amount of effort needed for them to perform well. However, treatment coefficients are never significantly negative, implying that, even for low-ability students, the negative effects that financial incentives might produce on "intrinsic motivation"—related to the desire to reach

TABLE 4
The Impact of the Incentives on Credits Earned: OLS Regressions

	All (1)	Top (2)	Bottom (3)	Males (4)	Females (5)
Treatment A (€700)	2.335*	4.156**	.533	1.759	2.490
	(1.217)	(1.791)	(1.628)	(1.854)	(1.588)
Treatment B (€250)	2.194*	4.225**	.076	.714	2.766*
	(1.266)	(1.774)	(1.744)	(1.970)	(1.655)
Female	−.245	−2.348	1.267		
	(1.069)	(1.575)	(1.372)		
Age	−.554***	−2.079**	−.486**	−.189	−.476
	(.205)	(.948)	(.189)	(.308)	(.453)
High school grade	.471***	1.348***	.181*	.453***	.517***
	(.082)	(.311)	(.097)	(.120)	(.107)
Scientific lyceum	12.566***	7.910**	2.206	15.387***	10.367***
	(1.408)	(3.394)	(4.201)	(2.108)	(1.972)
Classical lyceum	5.393**	.000	2.074	−1.173	5.164
	(2.681)	(.000)	(4.589)	(5.546)	(3.158)
Technical schools	6.194***	−4.714	3.021**	9.293***	3.304
	(1.393)	(4.189)	(1.500)	(1.987)	(2.013)
Class 1: A–L	.749	2.428*	−1.202	−.737	2.005
	(1.035)	(1.434)	(1.455)	(1.700)	(1.334)
Lyceum × high school grade	.251**	−.531	−.163	.261	.240
	(.112)	(.327)	(.294)	(.185)	(.155)
Constant	−21.962***	−65.010**	4.007	−29.225**	−26.257*
	(8.235)	(30.433)	(9.211)	(12.181)	(14.867)
Observations	462	228	234	206	256
R^2	.314	.244	.136	.376	.301

Note.—The dependent variable is credits, the number of credits earned. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions we control for province of residence dummies (six categories, not reported).
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

the objective for itself rather than for a financial reward—are not relevant or not strong enough to counterbalance the positive effects that these incentives produce on "extrinsic motivation."

In columns 4 and 5, we report separate estimates for males and females, respectively. The effects of the incentives seem to be more pronounced for females, but the difference between males and females turns out to be statistically insignificant when we estimate the regression in column 1 using two interaction terms between the treatment status and the dummy female (not reported).

In table 4, estimates are reported by considering the number of credits earned by students as a dependent variable. Results are very similar to those obtained when measuring student performance using total points. The average treatment effects of both the large and the small financial rewards are positive and statistically significant (at the 5 and 10 percent levels, respectively). The treatment effects are equal to about 0.17–0.18 standard deviations of the dependent variable. The magnitude of the two treatment effects is quite similar, implying that the size of the prize is not particularly relevant for the average student in the population. These findings appear to contrast with the results of Gneezy and Rustichini (2000), who find that small monetary rewards tend to be worse than no reward at all.

Results in line with those discussed above also emerge when we estimate separately according to student predicted ability; we find that students with higher predicted ability react strongly to both the large and the small rewards (earning about four credits more than those in the control group). Finally, as shown in columns 4 and 5, females seem more reactive to financial incentives than males, but again differences are not statistically significant.

To further investigate the differences existing between high- and low-ability students, we consider the high school grade instead of using the student predicted performance as a measure of ability. In table 5, the effects are reported separately for students with ability above and below the median high school grade of 90. In columns 1 and 2, the dependent variable is total points, whereas in columns 3 and 4, student performance is measured considering credits. It emerges that the effect of financial rewards on total points is negligible for lower-ability students, whereas both the small and the large rewards produce a positive effect in higher-ability students.

Very similar results emerge when we measure student performance using credits: students with higher ability react strongly to the large reward (earning 3.8 credits more than the control group), whereas the effect of the small reward is weaker.

IV. The Effects of Incentives by Using Course-Level Observations

In this section, we stack data with the aim of using student-course-level observations (462 students × 6 courses = 2,772) to estimate the treatment effects on the student's probability of passing an exam and on the grade obtained at each exam.

In table 6, we report estimates of a linear probability model in which the dependent variable, pass, is a dummy equal to one when the student passes a given exam, and zero otherwise. The mean value of pass is 0.38, and its standard deviation is 0.48. In addition to the usual individual characteristics, we control for course dummies (calculus is the reference category) to take into account heterogeneous levels of difficulty among courses and/or different grading standards among instructors. Standard errors are robust to heteroskedasticity and are clustered at the student level.

Students in the treatment groups show a higher probability of passing exams: the incentives generate an increase of about 5 percentage points. The effect is statistically significant at the 5 and 10 percent levels for the large and the small rewards, respectively.

TABLE 5
Differences between High- and Low-Ability Students (Distinguished by High School Grade): OLS Regressions

	Total Points		Credits	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)
Treatment A (€700)	9.488**	.408	3.827**	.065
	(4.710)	(3.612)	(1.833)	(1.532)
Treatment B (€250)	5.889	4.017	2.529	1.616
	(4.608)	(4.118)	(1.804)	(1.733)
Observations	236	226	236	226
R^2	.356	.252	.346	.246

Note.—Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions we control for gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence, and an interaction term between high school grade and lyceum.
** Statistically significant at the 5 percent level.

When we run separate regressions for students of different predicted abilities,[14] it emerges that only students with a level of predicted ability above the average react to the financial incentives (col. 2), whereas incentives do not have any impact on low-ability students (col. 3). The effects of the large and the small rewards are almost the same. Again, we do not find any statistically significant difference between males and females (not reported). Very similar results are obtained when using a probit estimator instead of a linear probability model.

In table 7, we investigate the effects of financial incentives on the grades obtained by students at exams using the same specifications as in table 6. Results from OLS estimates suggest that the incentives are effective not only in enhancing students' probability of passing exams but also in increasing the grades they obtain. Being included in the treated groups increases student grades by about 2 points (statistically significant at the 5 percent level). Once again, treatment effects turn out to be significant for students with high predicted ability but not for students with low predicted ability.

In the estimates shown in table 7, we have also considered, when possible, grades obtained in failed examinations (below the minimum pass threshold of 18/30); for those examinations for which grades below this minimum pass mark were not recorded (public law and business administration), the grade was set equal to zero. As an alternative strategy, we have considered all grades below the minimum pass mark as censored, and we have estimated a Tobit model considering 17 as the lower limit for left censoring and 31 as the upper limit for right censoring. Results (not reported) are quite similar to those presented in

[14] We follow the same procedure explained in Sec. III and consider the predicted performance of students in terms of the probability of passing exams.

TABLE 6
INCENTIVES AND THE PROBABILITY OF PASSING AN EXAM: LINEAR PROBABILITY MODEL

	All (1)	Top (2)	Bottom (3)
Treatment A (€700)	.051**	.094***	.009
	(.025)	(.036)	(.033)
Treatment B (€250)	.048*	.093**	.001
	(.026)	(.036)	(.035)
Accounting	.500***	.420***	.580***
	(.024)	(.034)	(.034)
Microeconomics	.043*	.030	.056**
	(.023)	(.036)	(.028)
Statistics	−.251***	−.407***	−.095***
	(.021)	(.034)	(.022)
Business administration	.063**	.022	.104***
	(.027)	(.045)	(.032)
Public law	.069**	.013	.126***
	(.028)	(.046)	(.033)
Female	−.003	−.041	.024
	(.022)	(.031)	(.027)
Age	−.012***	−.037*	−.010**
	(.004)	(.019)	(.004)
High school grade	.010***	.022***	.002
	(.002)	(.007)	(.002)
Technical schools	.130***	−.064	.042
	(.028)	(.089)	(.030)
Scientific lyceum	.254***	.129*	.112
	(.029)	(.068)	(.079)
Classical lyceum	.133**	.000	.073
	(.056)	(.000)	(.095)
Class 1: A–L	.006	.037	−.035
	(.021)	(.028)	(.030)
Lyceum × high school grade	.005**	−.005	.001
	(.002)	(.007)	(.006)
Constant	−.519***	−.942	.082
	(.166)	(.664)	(.184)
Observations	2,772	1,386	1,386
R^2	.298	.279	.283

Note.—The dependent variable is the dummy pass. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the student level. In all the regressions, we control for province of residence dummies (six categories, not reported).
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

table 7. In particular, on the basis of the Tobit estimations, we are able to determine the marginal effect of the treatment on the probability that a student has passed an exam (i.e., the effect on the probability that the grade is greater than 17): this effect is equal to 0.06. Moreover, given that a student has passed an exam (i.e., he or she is not censored), we can calculate the marginal effect of the treatment on the expected grade: being included in the large-reward group increases the expected grade by 0.29.

TABLE 7
INCENTIVES AND GRADES OBTAINED AT EXAMINATIONS: OLS REGRESSIONS

	All (1)	Top (2)	Bottom (3)
Treatment A (€700)	1.327**	2.058**	.315
	(.597)	(.865)	(.824)
Treatment B (€250)	.930	2.447***	.161
	(.630)	(.909)	(.835)
Accounting	11.227***	10.707***	12.839***
	(.420)	(.730)	(.708)
Microeconomics	4.221***	1.746**	1.661**
	(.407)	(.759)	(.643)
Statistics	−5.719***	−8.672***	−1.817***
	(.428)	(.747)	(.457)
Business administration	.268	1.185	3.357***
	(.591)	(1.020)	(.716)
Public law	−.199	.526	3.083***
	(.586)	(1.031)	(.690)
Female	.488	−1.412*	1.120
	(.519)	(.795)	(.679)
Age	−.350***	−.900*	−.235***
	(.103)	(.513)	(.086)
High school grade	.223***	.591***	.094**
	(.039)	(.151)	(.045)
Technical schools	3.174***	.000	1.195
	(.672)	(.000)	(.724)
Scientific lyceum	6.012***	4.874***	4.012**
	(.713)	(1.248)	(1.694)
Classical lyceum	3.063**	1.960	1.349
	(1.364)	(1.895)	(2.031)
Class 1: A–L	−.344	1.069	−.988
	(.512)	(.724)	(.707)
Lyceum × high school grade	.140**	−.113	.063
	(.059)	(.163)	(.114)
Constant	−8.757**	−30.216*	−1.581
	(4.124)	(16.366)	(4.776)
Observations	2,772	1,392	1,380
R^2	.330	.286	.255

Note.—The dependent variable is grade. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the student level. In all the regressions, we control for province of residence dummies (six categories, not reported).
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

V. Dealing with Partial Compliance by Using an Instrumental Variable Estimation Strategy

In the previous sections, we have analyzed intention-to-treat effects, since we have considered the whole population of students enrolled in the first year of the business administration degree course, although not all students signed up to show their interest and awareness of the experiment. Therefore, the estimated effects are diluted by the fact that some treated students may actually not have been informed of the experiment

TABLE 8
Two-Stage Least-Squares Estimates of the Effects of Effective Participation to the Treatment

	Total Points			Credits		
	All (1)	Top (2)	Bottom (3)	All (4)	Top (5)	Bottom (6)
Treatment A effective	6.610**	11.355**	1.644	2.562*	4.379**	.611
	(3.335)	(4.956)	(4.355)	(1.327)	(1.863)	(1.857)
Treatment B effective	6.029*	11.142**	−.115	2.473*	4.419**	.093
	(3.545)	(4.699)	(5.117)	(1.417)	(1.833)	(2.148)
Female	−.831	−6.074	3.469	−.368	−2.340	1.254
	(2.611)	(3.983)	(3.234)	(1.054)	(1.543)	(1.401)
Age	−1.248**	−5.199**	−1.083**	−.565***	−2.071**	−.490**
	(.515)	(2.611)	(.447)	(.205)	(.943)	(.189)
High school grade	1.184***	3.201***	.381*	.455***	1.317***	.177*
	(.206)	(.693)	(.227)	(.083)	(.305)	(.098)
Scientific lyceum	31.372***	29.054***	7.109	12.386***	12.463***	2.124
	(3.498)	(5.675)	(8.820)	(1.388)	(2.524)	(4.163)
Classical lyceum	12.844**	6.086	6.753	5.159*	4.359	2.032
	(6.491)	(9.510)	(11.102)	(2.666)	(4.151)	(4.599)
Technical schools	14.909***	.000	5.941	5.982***	.000	2.999*
	(3.380)	(.000)	(3.702)	(1.387)	(.000)	(1.525)
Class 1: A–L	1.745	4.674	−2.193	.795	2.376*	−1.186
	(2.567)	(3.748)	(3.514)	(1.027)	(1.410)	(1.458)
Lyceum × high school grade	.864***	−.740	−.136	.259**	−.501	−.168
	(.290)	(.763)	(.599)	(.111)	(.319)	(.289)
Constant	−60.419***	−158.488**	10.305	−20.103**	−66.763**	4.378
	(20.231)	(76.309)	(21.396)	(8.126)	(31.690)	(9.051)
Observations	462	231	231	462	228	234
R^2	.340	.274	.138	.329	.267	.139

Note.—In the first stage, dummies for the assigned treatment are used as instruments for treatment A effective and treatment B effective. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions we control for province of residence dummies (six categories, not reported).
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

or may not be interested in it (see Bloom 1984; Angrist and Pischke 2009).

In this section, we analyze the impact of the financial incentives on the students who have effectively participated in the treatment, that is, those students who filled in the participation form. To deal with endogeneity problems that may affect participation, we instrument the effective participation in each group with the two randomly assigned treatment statuses (treatment A and treatment B)

Two-stage least-squares estimates using total points (cols. 1–3) and credits (cols. 4–6) as dependent variables are reported in table 8. Results show that adjusting intention-to-treat effects for nonparticipation leads to a stronger impact of the treatment on student performance, measured either as total points or as credits. The increase in the effect is proportional to the reciprocal of the participation rate.

The average treatment effect of the €700 reward is an increase of 14 percent in total points (with respect to the sample mean), and the effect of the small reward is an increase of 13 percent. The effects are higher when focusing on students with high predicted ability: when included in the treated groups, these students obtain about 11 points more, that is, an increase of 19 percent with respect to the average total points realized by students of comparable ability.

The same analysis has been conducted, using student-course-level observations, on the probability of passing examinations and for grades obtained. Results are similar to those discussed above (not reported).

VI. Some Evidence from a Postexperiment Survey

The results shown in our analysis might be related, in addition to the financial incentives offered in our project, to a number of other factors taking place simultaneously with the experiment and correlated to the treatment. In order to investigate these aspects, we have conducted a postexperiment survey among our sample students. About 80 percent of students answered the questionnaire.

First, peer effects or other externalities deriving from the possibility that better-performing students may influence the performance of students in the control group positively and potentially produce a downward bias in our estimation results. To try to tackle this issue, we asked students the names of the other students (maximum four) they study with most often. Then, we built a variable, treated peers, measuring the number of treated students among each student's peers (without distinguishing between treatments A and B).

Peers are not exogenously determined since individuals typically choose people to associate with. In principle, students might select their peers according to their treatment status; for example, students assigned to the treatment groups may develop studying habits that induce them to meet more often (e.g., in the library). To check whether students selected their peers according to their treatment status, we regressed the number of treated peers of each student on the treatment status of the student. We do not find any statistically significant effect (p-value = .751), suggesting that peers are not chosen according to their treatment status.

Therefore, the experiment introduced an exogenous change in the study effort of peers, which may also have induced students in the control group to spend more time in studying activities. With the aim of uncovering this effect, we ran a regression of both total points and credits on treated peers and the usual controls (as in table 3, col. 1). The estimation results (not reported) seem to show that the number of treated peers does not affect student performance. The coefficient is far from being statistically significant (p-value = .79). However, these results are only suggestive since students participating in the survey might not be a random sample of the population of interest.

We have also checked whether students assigned to the treatment groups are more likely to choose high-ability peers. We find that students eligible for the monetary incentives tend to choose high-ability peers (measured with their high school grade). Although the effect is weakly statistically significant (p-value = .15), we cannot exclude that the total effect of the monetary reward incorporates a spillover effect arising from being with higher-ability peers.

Second, following Leuven et al. (2010), to be sure that the treatment in our experiment has not been modified by the intervention of other subjects, we asked students whether parents or other subjects promised additional rewards to those established by the experiment. From students' answers it clearly emerges that supplementary rewards were rather uncommon. Only 2 percent of students report that their parents promised them a monetary prize in the case of good performance in their first-year examinations.

Third, it could be the case that the effect we have found was negatively influenced by poor cooperation among peers induced by the design of the experiment. The experiment had a tournament structure, which, as is well known in the agency literature (see Prendergast 1999), tends to discourage cooperation. In our postexperiment survey, we asked students whether the experiment had led them to be less cooperative with their peers and whether or not they had obtained cooperation in terms of course materials, information sharing, and support in study activities from their colleagues. About 3 percent of students admitted to having been much less cooperative with their peers as a result of the experiment, whereas about 15 percent of students said that their colleagues had not been cooperative in studying activities. Students included in the treated groups are both less inclined to offer cooperation to and less likely to receive cooperation from their peers, supporting the idea that the tournament structure of the experiment may have, at least in part, reduced cooperation. Therefore, it is likely that the treatment effects we find would have been larger if we had used a reward system based on an absolute, rather than on a relative, measure of performance.

VII. Higher Effort or Substitution Effects?

Binder, Ganderton, and Hutchens (2002) and Cornwell, Lee, and Mustard (2005) find that scholarships are effective in increasing student grades but also induce students to take less ambitious courses. In contrast to these works, our findings show that financial rewards do not lead to adverse changes in student behavior. Our sample students have both improved their grades and increased the credits earned, suggesting that the improvement in their academic performance can be attributed to a higher level of effort rather than to the attempt of "gaming" the incentive system. Given the experimental design adopted, where we define the target both in terms of grades and in terms of the number

TABLE 9
Probability of Passing the Exams Not Included in the Target Variable: Linear Probability Model

	All (1)	Top (2)	Bottom (3)	Males (4)	Females (5)
Treatment A (€700)	.043	.007	.073	.053	.034
	(.032)	(.038)	(.052)	(.053)	(.038)
Treatment B (€250)	−.026	.013	−.076	−.062	−.010
	(.034)	(.041)	(.055)	(.053)	(.044)
Observations	1,386	699	687	618	768
R^2	.246	.125	.278	.275	.209

Note.—The dependent variable is pass in English, French, and computer sciences examinations. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the student level. In all the regressions, we control for gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence, and an interaction term between high school grade and lyceum. Top and bottom students are defined according to their predicted performance measured through total points.

of credits earned, this is, to some degree, an expected result. Students had little chance of obtaining the monetary rewards by simply substituting effort from nontargeted tasks to targeted ones so as to leave their effort levels unchanged.

To further investigate this issue, we analyze student pass rates in a number of examinations that were not defined as targets for the experiment. As mentioned in Section II, students enrolled in the first year of the degree in business administration are required to take some additional examinations (English, French, and computer sciences), which were not included in the target variable total points. The decision to exclude these exams was made because no grade is awarded for them and only a pass or fail result is observed.[15] The average probability of passing these examinations is .70.

Table 9 reports linear probability model estimates of student pass rates in these examinations by using student-course-level observations ($462 \times 3 = 1{,}386$ observations). We do not observe any significant difference between students in the treated and the control groups with regard to their performance in nontarget examinations.

This result has two important implications. First, without financial incentives, students in the treated groups do not outperform those in the control group, confirming the effectiveness of monetary incentives in enhancing student performance. Second, it suggests that "multitasking" problems are not a major concern in our experiment: students eligible for the financial rewards—while increasing effort to improve their "measurable" performance—did not reduce their effort in other study activities that were not directly relevant in obtaining the prize.

[15] Moreover, students may have very different initial levels of unobservable competence in these specific subjects.

TABLE 10
THE IMPACT OF FINANCIAL INCENTIVES ON CLASS ATTENDANCE: OLS REGRESSIONS

	All (1)	Top (2)	Bottom (3)
Treatment A (€700)	.162	.201	.154
	(.149)	(.221)	(.201)
Treatment B (€250)	−.069	.058	−.233
	(.145)	(.213)	(.203)
Observations	462	231	231
R^2	.144	.145	.139

Note.—The dependent variable is attendance. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions, we control for gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence, and an interaction term between high school grade and lyceum. Top and bottom students are defined according to their predicted performance measured through total points.

Therefore, we can conclude that the improvement in student achievements discussed in the previous sections was determined by a higher provision of effort in studying activities.

We have also tried to assess whether the increased level of student effort has taken the form of higher attendance in teaching classes. Students can improve their performance by devoting more time to studying activity at home, attending a higher number of hours in classes, or participating more regularly in tutorials with teaching assistants. We do not have information regarding time spent in study activities, but we have collected data on student attendance. As checking students' attendance was very time demanding,[16] in order not to delay the teaching schedule, we checked each course just once through an unannounced check on a randomly chosen day (six checks in total). On average, students were found in class 2.75 times out of six, implying that they attended about half of the classes.

We have used these data to build a variable, attendance, equal to the number of times the student was found attending a class. As shown by OLS estimates, reported in table 10, where attendance is regressed on the treatment dummies and on the usual controls, the financial reward does not significantly increase student presence in classes. We find a positive, weakly significant effect (p-value = .25) for treatment A and a null effect for treatment B. These findings suggest that the better performance of treated students is not related to higher class attendance.

[16] To avoid cheating, instead of simply asking students to sign or mark a paper with their name, we have ascertained their identity by checking an identification document.

TABLE 11
THE IMPACT ON STUDENTS' PERFORMANCE IN SUBSEQUENT YEARS (Credits): OLS REGRESSIONS

	Following 2 Years (1)	Second Year (2)	Third Year (3)
Treatment A (€700)	6.510	3.018	3.492
	(4.413)	(2.098)	(2.636)
Treatment B (€250)	8.058*	3.482*	4.575*
	(4.517)	(2.170)	(2.685)
Observations	462	462	462
R^2	.287	.267	.256

Note.—The dependent variable is credits acquired, respectively, in the second and third years. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions we control for gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence, and an interaction term between high school grade and lyceum.
* Statistically significant at the 10 percent level.

VIII. Effects on Students' Subsequent Performance

An important question in evaluating the effects of financial incentives is that of the effect on students' behavior once the incentives are removed. It could be the case that students who are given financial rewards for a good academic performance feel less engaged once the incentives are withdrawn and end up obtaining worse results than they would have done in the absence of the program (crowding-out effect). On the other hand, financial incentives may establish good habits and also produce positive effects when they are no longer available.

Only a few works have investigated this aspect, providing a mixed picture. Bettinger (2010) finds no evidence that the positive effects of financial incentives persist into subsequent years when students are no longer eligible for cash payments. Angrist et al. (2009) find no effect of financial incentives 1 year after their withdrawal, whereas a positive effect emerges for students who had both financial incentives and academic support services. Leuven et al. (2010) show that high-ability students tend to benefit from financial incentives obtained during their first year also in subsequent years, whereas low-ability students end up obtaining worse results also after the financial scheme has been removed.

To investigate this issue we have collected data on students' academic performances during the second and third years of their degree program (the 2 years following the withdrawal of the incentive scheme). The results obtained when considering as dependent variables the number of credits gained by students in the 2 years after the program's removal (col. 1) and the number of credits obtained during the second year (col. 2) and the third year (col. 3), respectively, are reported in table 11.

It emerges from column 1 that students assigned to the large reward

TABLE 12
Differences between High- and Low-Ability Students in Subsequent Performance: OLS Regressions

	Following 2 Years (1)		Second Year (2)		Third Year (3)	
	Top	Bottom	Top	Bottom	Top	Bottom
Treatment A (€700)	8.538	4.412	4.617	1.629	3.920	2.783
	(6.189)	(6.251)	(2.978)	(2.965)	(3.756)	(3.705)
Treatment B (€250)	13.700**	1.131	6.717**	−.062	6.984*	1.193
	(6.222)	(6.608)	(2.929)	(3.243)	(3.862)	(3.798)
Observations	231	231	231	231	231	231
R^2	.157	.177	.145	.176	.139	.155

Note.—The dependent variable is credits acquired, respectively, in the second and third years. Standard errors (reported in parentheses) are corrected for heteroskedasticity. In all the regressions, we control for gender, age, high school grade, type of high school dummies, assigned teaching class, dummies for province of residence, and an interaction term between high school grade and lyceum. Top and bottom students are defined according to their predicted performance measured through total points.
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

gain six credits more in the 2 years following the experiment compared to students assigned to the control group (weakly significant with a p-value of .14). The assignment to the small-reward group produces an increase in student performance of about eight credits, statistically significant at the 10 percent level.

In the other two columns, where we consider student performance in the second and third years separately, it emerges that the positive spillovers of the experiment persist into each of the subsequent years. The effect is positive and statistically significant for students assigned to the small-reward group, whereas a positive but more imprecisely estimated effect is found for students assigned to the large-reward group.

In table 12, we give the estimation results obtained when considering low- and high-ability students separately, considering predicted performance as a measure of ability. The long-run positive effects are concentrated on high-ability students who were assigned to the small-reward group. When the financial incentives were removed, high-ability students who were assigned to treatment B continued to obtain a higher and statistically significant number of credits than students in the control group. A positive, but imprecisely estimated, effect emerges for students assigned to the large-reward group. In addition, while the positive effect of being assigned to the large-reward group tends to persist only for the first year after the removal of the program, being assigned to the small-reward group also produces positive effects 2 years later.

For treated high-ability students, the incentive scheme seems to produce positive effects not only during the period in which incentives are in place but also in subsequent periods. This effect could be due either to the fact that students producing a good performance during the first year accumulate skills that are useful for more advanced courses or to the fact that they have acquired better studying habits. Instead, we do not find any statistically significant effect of treatment for low-ability students in the 2 years after the experiment. Although financial rewards do not have any beneficial effects for these students, it is also true that they do not seem to undermine the intrinsic motivation of low-ability students (who are probably more susceptible to crowding-out effects) sufficiently to lead to negative outcomes.

IX. Concluding Remarks

Merit scholarships may constitute an incentive for students to provide more effort, and they may help them to improve their attitudes toward study activities. This is especially true in those labor markets in which there are a number of distortions that lead to relatively low expected returns on human capital investments and discourage students from providing high effort in study activities.

The empirical investigations that try to shed some light on the effects of financial incentives for students have reached heterogeneous results, suggesting the need to extend the research in order to better understand which factors are most likely to affect the success of this type of policy.

In this paper, we have provided some additional evidence on the effects of financial rewards on student academic performance, presenting the results of an experiment conducted among a sample of undergraduate students enrolled at a southern Italian public university. It emerges from our analysis that financial incentives significantly improve student performance. The students included in the treated groups obtained better grades and a higher number of credits. In terms of standard deviations, student performance increased by about 0.18. Compared to other educational policies, the cost of the project—a total outlay of €28,500 in prizes for students—also appears quite small. Performing a "back-of-the-envelope" calculation and considering the reduction of class size as an alternative strategy for increasing student performance, it emerges that monetary incentives are a cost-effective policy. In fact, if we take for granted the estimates of Bandiera, Larcinese, and Rasul (2010), which indicate that a reduction in class size from 211 to 104 increases college student performance by about 0.17 standard deviations (see table 4 in their paper), and suppose that these estimates are valid for our sample of students, a reduction in class size from 230 (the current size) to 115 (by organizing two additional teaching classes) would have almost the same effect on performance as the monetary incentives used in our experiment. However, the cost of organizing two additional teaching classes to increase the number of classes from the current two to four, simply considering the additional expense of instructors' remunerations, would imply a cost of at least €80,000, a figure well in excess of the cost of the incentives given to students.

The cost might have been even smaller as we find that large and small rewards produce very similar effects, implying that the size of the prize is not particularly relevant in enhancing student effort. Interestingly, a small financial reward seems sufficient to boost student performance not only for the period during which the incentives are in place but also after their removal.

In line with the findings emerging from other works, we find that both short- and long-run benefits from the treatment are mainly concentrated among students with higher ability. When included in the treated groups, these students perform significantly better than students of comparable ability in the control group, whereas the effect of treatment on low-ability students is not statistically significant. Our evidence suggests that even though our sample students were optimistic about their academic capabilities, the definition of a particularly demanding target leads many of them to abandon the competition. Therefore, if the policy maker aims to improve the performance of low-ability students, a different incentive scheme should be adopted (either restricting competition to low-ability students or taking into account improvements in performance with respect to the students' initial abilities). On the other hand, the policy design implemented in our experiment seems able to increase the performance of more promising students, which may represent an important resource in depressed economies in trying to create a new, more competent ruling class, so increasing the possibility of success in fostering growth.

In contrast to other studies that find that financial incentives may lead to adverse changes in student behavior, we do not find evidence of any dysfunctional responses. The improvement in student performance detected in our research seems to be imputable to a higher level of effort in study activities: students in the treatment groups obtained better grades and earned a higher number of credits without obtaining worse results in those examinations that were excluded from the experiment. Furthermore, we do not find any crowding-out effect of monetary incentives in the short run (when incentives are in place) or in the long run (when incentives are removed). In the long run, treated students continue to perform better than control students, implying that incentives generate dynamic positive externalities. Moreover, low-ability students do not show any negative reaction, implying that the negative effects that financial incentives may produce on intrinsic motivation are not particularly relevant.

These positive findings may be related to the economic and social context in which our sample students live: in depressed labor markets, such as those characterizing the South of Italy, students may be more reactive to financial incentives since they lack the incentives deriving from a well-functioning labor market that offers high rewards to skills in terms of better employment opportunities and higher wages.

Appendix

A Very Simple Model: Student Effort and Incentives

In this appendix, we present a very simple model of student effort and incentives to show that financial rewards may be more effective in enhancing student performance in environments characterized by low returns to skills on the labor market.

For the sake of simplicity, let us assume that the educational process lasts only a period. At the end of that period, the individuals who meet a certain standard obtain both a wage return in the labor market and a financial incentive, whereas those not attaining the standard will obtain zero (see, e.g., Costrell 1994).

Individuals have the following lifetime utility function:

$$V = P(e)U(W_K + M) - c(e),$$

where e is the student effort; $P(e)$ is the probability of meeting the standard, with $P'(e) > 0$ and $P''(e) \leq 0$; $U(W_K + M)$ is the monetary utility function ($U' > 0$ and $U'' < 0$); W_K is the discounted flow of future wages in the labor market K (a measure of returns to skills); M is a financial incentive obtained on the basis of the academic performance; and $c(e)$ is the effort cost function with $c'(e) > 0$ and $c''(e) > 0$.

Maximizing V with respect to the student effort, e, we obtain the following first-order condition:

$$\frac{\partial V}{\partial e} = F = P'(e)U(W_K + M) - c'(e) = 0.$$

With the implicit function theorem (we call F the first-order condition), the reaction of student effort to monetary incentives can be obtained as

$$\frac{\partial e}{\partial M} = -\frac{\partial F/\partial M}{\partial F/\partial e},$$

where

$$\frac{\partial F}{\partial M} = P'(e)U'(W_K + M)$$

and

$$\frac{\partial F}{\partial e} = P''(e)U(W_K + M) - c''(e)$$

(note that the latter is the second-order condition, which must be negative for a maximum). Therefore,

$$\frac{\partial e}{\partial M} = -\frac{P'(e)U'(W_K + M)}{P''(e)U(W_K + M) - c''(e)},$$

which is positive given that $P' > 0$, $U' > 0$, and the second-order condition is negative.

We are interested in understanding how the student effort reacts to monetary incentives in relation to changes in W_K. Therefore, taking the derivative with respect to W_K, we obtain

$$\frac{\partial e/\partial M}{\partial W_K} = -\frac{[P''(e)U(W_K + M) - c''(e)][P'(e)U''(W_K + M)] - [P'(e)U'(W_K + M)]P''(e)U'(W_K + M)}{[P'(e)U(W_K + M) - c''(e)]^2}.$$

On the basis of the above assumptions, we can easily sign the terms of the expression as follows:

$$\frac{\partial e/\partial M}{\partial W_K} = -\frac{\overbrace{[P''(e)U(W_K + M) - c''(e)]}^{-}\ \overbrace{[P'(e)U''(W_K + M)]}^{-} - \overbrace{[P'(e)U'(W_K + M)]}^{+}\ \overbrace{P''(e)U'(W_K + M)}^{-}}{\underbrace{[P'(e)U(W_K + M) - c''(e)]^2}_{+}} < 0.$$

Therefore, we have shown that $\partial e/\partial M$ is higher in labor markets with lower returns to skills.

References

Alesina, A., S. Danninger, and M. Rostagno. 2001. "Redistribution through Public Employment: The Case of Italy." *IMF Staff Papers* 48 (3): 447–73.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *A.E.R.* 92 (5): 1535–58.

Angrist, J., E. Bettinger, and M. Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *A.E.R.* 96 (3): 847–62.

Angrist, J., D. Lang, and P. Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Econ. J.: Appl. Econ.* 1 (1): 136–63.

Angrist, J., and V. Lavy. 2009. "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." *A.E.R.* 99 (4): 1384–1414.

Angrist, J., and S. Pischke. 2009. *Mostly Harmless Econometrics.* Princeton, NJ: Princeton Univ. Press.

Arcidiacono, P., V. J. Hotz, and S. Kang. 2012. "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals." *J. Econometrics* 166 (1): 3–16.

Bandiera, O., V. Larcinese, and I. Rasul. 2010. "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students." *Econ. J.* 120:1365–98.

Bettinger, E. 2010. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." Working Paper no. 16333, NBER, Cambridge, MA.

Binder, M., P. T. Ganderton, and K. Hutchens. 2002. "Incentive Effects of New Mexico's Merit-Based State Scholarship Program: Who Responds and How?" In *Who Should We Help? The Negative Social Consequences of Merit Scholarships*, edited by D. E. Heller and P. Marin, 43–56. Cambridge, MA: Harvard Univ. Press.

Bloom, H. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Rev.* 8:225–46.

Cornwell, C., K. Lee, and D. Mustard. 2005. "Student Responses to Merit Scholarship Retention Rules." *J. Human Resources* 40 (4): 895–917.
Costrell, R. 1994. "A Simple Model of Educational Standards." *A.E.R.* 84 (4): 956–71.
Croson, R., and U. Gneezy. 2009. "Gender Differences in Preferences." *J. Econ. Literature* 47 (2): 448–74.
De Paola, M., and V. Scoppa. 2007. "Returns to Skills, Incentives to Study and Optimal Educational Standards." *J. Econ.* 92 (3): 229–62.
Fabbri, F., and N. Rossi. 1997. "Caste, non classi: Una società immobile." *Il Mulino* 1:110–16.
Fryer, R. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Q.J.E.* 126:1755–98.
Gneezy, U., S. Meier, and P. Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *J. Econ. Perspectives* 25 (4): 191–210.
Gneezy, U., and A. Rustichini. 2000. "Pay Enough or Don't Pay at All." *Q.J.E.* 115 (3): 791–810.
Heller, H. 2006. "Merit Aid and College Access." Paper presented at the Symposium on the Consequences of Merit-Based Student Aid, Wisconsin Center for the Advancement of Postsecondary Education, Madison, January 3.
Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design." *J. Law, Econ., and Org.* 7:24–52.
Kremer, M., E. Miguel, and R. Thornton. 2009. "Incentives to Learn." *Rev. Econ. and Statis.* 91 (3): 437–56.
Leuven, E., H. Oosterbeek, and B. van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." *J. European Econ. Assoc.* 8 (December): 1243–65.
Levitt, S., J. List, and S. Sadoff. 2010. "The Impact of Short-Term Financial Incentives on Student Performance." Manuscript, Univ. Chicago.
Prendergast, C. 1999. "The Provision of Incentives in Firms." *J. Econ. Literature* 37 (1): 7–63.
Rodriguez-Planas, N. 2010. "Mentoring, Educational Services, and Economic Incentives: Longer-Term Evidence on Risky Behaviors from a Randomized Trial." IZA Discussion Paper no. 4968, Inst. Study Labor, Bonn.
Schultz, T. P. 2004. "School Subsidies for the Poor: Evaluating Mexico's Progresa Poverty Program." *J. Development Econ.* 37:199–250.
Stinebrickner, R., and T. Stinebrickner. 2009. "Learning about Academic Ability and the College Drop-Out Decision." Working Paper no. 14810, NBER, Cambridge, MA.