Joint Liability vs. Individual Incentives in the Classroom. Lessons from a Field Experiment with Undergraduate Students

Cid, Alejandro and Cabrera, José María

Universidad de Montevideo

2012

Joint Liability vs. Individual Incentives in the Classroom.

Lessons from a Field Experiment with Undergraduate Students

José María Cabrera[1]
Alejandro Cid[2]
July, 06th. 2012.

Center for Research on Applied Economics
Universidad de Montevideo

Abstract

We evaluate the impact of joint-liability incentives in the classroom using a randomized field experiment. The instructor design groups of three students in the classroom and provides a premium to their homework's grade only if all three members of the group accomplish some requirements. To isolate the joint liability effect from selfish motivations, we design also an individual incentives treatment. We find that joint-liability incentives impact positively on the grades accomplished in homework and midterm exams both in the experimental courses and in the other courses taken by the students in the semester. Though the positive average effect seems to disappear in the final exams, the overall impact of joint-liability incentives on the academic achievements in the semester is still positive. A drawback of this program is a decrease in the satisfaction with classmates. The significant effectiveness of the peer monitoring developed by the joint liability of group incentives provides novel implications for the design of the grading policies in the classroom and for other social settings where incentives may be based in peer monitoring or joint liability.

Keywords: field experiment; randomization; education; joint liability; student incentives

JEL: I20, I23

[1] jmcabrera@um.edu.uy
[2] acid@um.edu.uy The corresponding author. Prudencio de Pena 2544, Montevideo 11600, Uruguay, Tel. & Fax: +598 27074461.

We gratefully acknowledge the guidance of Martín Rossi. We also thank Ana Balsa, Marcelo Caffera, Juan Dubra, and Danilo Trupkin and seminar participants at Universidad de Montevideo for valuable comments. All errors remain our own.

I. Introduction

Incentives for teachers have received considerable attention in previous literature. Less attention has been paid to encouraging students (Angrist, Oreopoulos, and Williams, 2010; Angrist, Lang, and Oreopoulous 2009; Fryer 2010; Grant and Green, 2012) and it is not conclusive. For instance, recent research suggests that grades designed as individual incentives, or even monetary rewards, are not always effective motivators for students. Grading schemes have evolved with the history of educational system, partly in response to demands for better information about undergraduate performance, but were not explicitly designed to motivate students (Grant and Green, 2012).

We evaluate a novel design of incentives for students. That is, we design a joint liability contract that gives students strong incentives to monitoring each other. The instructor design groups of three students in the classroom and provides a premium to their homework's grade only if all three members of the group accomplish some requirements. To avoid self-virtuous group selection, we randomly assign participants to each group. And, in order to disentangle the pure effect of peer monitoring from the simple self motivation, we also randomly assign students to a group of individual incentives. Hence, employing randomization, we assign students to the joint-liability treatment, to the individual incentives treatment and to the control group.

The experimental courses are core ones for freshmen students at Universidad de Montevideo, a private university in Uruguay -a developing country of Latin America. The course composition is primarily undergraduate students majoring in economics, management and accountancy.

We find that joint-liability incentives impact positively on the grades accomplished in homework and midterm exams both in the experimental courses and in the other courses

taken by the students in the semester. Though the positive average effect seems to disappear in the final exams, the overall impact of joint-liability incentives on the academic achievements in the semester is still positive. On the other hand, the individual-incentive scheme has no effect. This result is in line with previous literature that provides not conclusive evidence about the effect of individual incentives on grades.

The significant effectiveness of the peer monitoring developed by the joint liability of group incentives provide novel implications for the design of the grading policies in the classroom and for other social settings where incentives may be based in peer monitoring or joint liability.

The rest of the paper is organized as follows. Section II describes the program and explains the experiment design. Section III presents the econometric model and the results. Section IV concludes.

II. Program and experiment design

Program

Undergraduate students at Universidad de Montevideo have to complete a number of credits in core courses in order to achieve their bachelor degree[3]. Two of these core courses are Macroeconomics I and Descriptive Economics, and students usually take them at their first year in the university. These two courses were structured in the same way in the 2011 academic year: a midterm exam (35% of the final grade), eight take-home-tests (15%), and a final exam (50%)[4]. The minimum acceptable grade to approve the course is 6 in a rank from 1 to 12. Also, attendance to classes is mandatory. Each course has sixty classes of fifty minutes each during fifteen weeks and students may have fifteen absences. There is nothing atypical

[3] Each course may have different credits. One credit corresponds to ten hours of classes.
[4] The frequency of take-home-tests is nearly one per two weeks. Instructors determined this number of take-home-tests looking for a sufficient number of occurrences that may form habit of exercising.

about these courses characteristics or their grading system in comparison with the other courses offered by the Universidad de Montevideo. We build a program that consists in giving incentives for take-home-tests and attendance.

We want to test if incentives designed as a joint liability scheme improves academic outcomes. We face two major challenges to identify this causal effect. The first one is *self virtuous group selection* (no one wants to mix with lazy classmates to minimize the probability of losing the reward), that we overcome it with the random assignment of participants to groups. A second challenge is that if faced with an incentive, an individual may exert more effort, *whether he is in a group or not*. If we only have the joint liability treatment and a control group we cannot uncover selfish motivations to get the prize from pure peer monitoring. So we build two different treatments in the classroom: individual and joint liability incentives, and a control group. With this design we think that we can identify the pure monitoring effect of peers.

Thus, we randomly distributed students in three groups[5].

In the *Joint Liability* group (Treated group 1), the student is randomly assigned to a group of three students and receives a 20% increase in the grade of each take-home-test if <u>each student of her group</u> fulfills two conditions: she obtains a grade of at least 6 in the take-home-test, and has no absences during the week in which the take-home-test must be handed in.

In the *Individual Incentive* group (Treated group 2), the student receives a 20% increase in the grade of each take-home-test if she obtains a grade of at least 6 in the take-home-test,

[5] The grading of these take-home-tests is done by research assistants that do not know the distribution of the students among the different treatments.

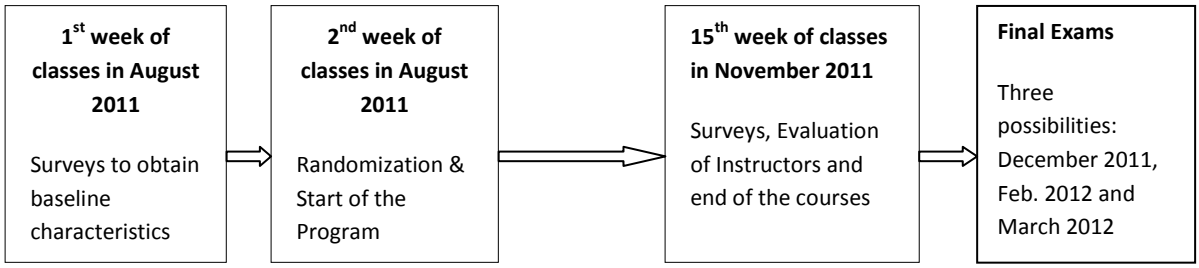
and has no absences that week. These are the same requirements as Treatment 1, but they don´t depend on the compliance of others.

In the *Control* group, the student does not receive any incentive besides the general conditions of grading in the course.

Take-home-tests in this field experiment do not require team work, even for students in treatment group 1. Each student is required to hand in his personal sheet with solutions at the beginning of the class and there's no problem if her solutions are identical to the ones of any other classmate.

For the evaluation design we use randomized trials, with the approval of the ethical review board of the university. There are 51 different students in this field experiment, 26 of them in Macroeconomics I and 25 in Descriptive Economics course. The selection process was as follows. In August 2011, all 51 applicants were subject to a survey. In this baseline survey we collected data on a wide array of students' characteristics such as age, gender, working hours, hours devoted to sports and volunteering, high school of origin and region of the country were they come from, distance between their home in Montevideo and the university, academic expectations and number of friends in the course. We have also administrative baseline data provided by the university such as grade average in previous courses and number of credits already completed at the university. From this population, given the restriction that the number of students in the *joint liability* group must be multiple of three, 24 students were randomly assigned to treated group 1, 14 to the treated group 2 and the remaining 13 candidates were assigned to the control group.

Timeline of the Program and Data Collection



Once the random allocation was performed, the balancing condition was checked. In case of significant differences at the ten percent level in mean pre-treatment characteristics between control and treated groups the random assignment procedure was repeated until we obtained an allocation that fulfills the balancing condition.

[Insert Table 1]

Table 1 reports the balancing condition and shows that the three groups have similar characteristics. They are balanced in eighteen observables variables. By the random allocation design, the probability of receiving a treatment is orthogonal to students characteristics, so including these characteristics in the regression model, while it may reduce standard errors, is not necessary for consistency[6].

As is normal in studies that follow students during period of classes, some observations suffered attrition. At November 2011 two individuals from Treatment Group 1, one from Treatment Group 2, and three of the Control Group dropped from the program. We have some outcomes for them during the courses and follow up administrative data, but we were not able to collect the complete data (grade at midterm exam, satisfaction with classmates or

[6] Our findings do not change if we include controls in the estimates.

evaluation of the instructor) in these six cases due to different reasons (most students are freshmen and usually a rate of them change to other degrees, some of them leave the course before the midterm exam and some refuse to evaluate the instructor because the evaluation demands extra time out of class).

We compare the pre-treatment characteristics between the individuals that have suffered attrition and those students who remain in the treated/control groups. Since fifteen from eighteen variables remain balanced, baseline data provide a measure of the similarity of these two groups. Only three variables are not balanced, that is, students that don't belong to Montevideo, students with fewer friends, and students with more unknown people in the class tend to drop more[7].

III. Econometric model and results

The primary purpose of this study is to determine the causal effect of Treatment 1 (joint-liability incentives to undergraduate students) and Treatment 2 (individual incentives) on students' achievements. Formally we want to estimate the following equation:

$$Y_i = a + bT1_i + cT2_i + dGroup_i + X_i'f + e_i \qquad (1)$$

where Y_i is one of the outcomes of interest for student *i* (*number of take-home-tests handed in, average grade at take-home-tests, grade of the midterm exam, grade at final exam, grade average in the midterm exams and homework of other simultaneous courses, average grade in the final exams of other simultaneous courses, accumulated grade average in the students career, total number of credits achieved in the semester*), $T1_i$ is a dummy variable that takes the value of one if student *i* is assigned to the treated group 1 and zero otherwise, $T2_i$ is a

[7] We include these variables in the regressions and our findings are not modified. They are available upon request.

dummy variable that takes the value of one if student *i* is assigned to the treated group 2 and zero otherwise, *b* and *c* are the parameters of interest, $Group_i$ is a dummy variable that takes the value of one if the student *i* belongs to the Macroeconomic course and zero otherwise, X_i is a matrix of students' characteristics, and e_i is the error term. Given that there's no problem of no-compliers, we can estimate this equation consistently with Ordinary Least Squares (OLS).

Prior research suggests that graded homework causes students to spend more effort relative to assigning non-graded homework (Pozo and Skull, 2006). Does providing joint-liability and individual extra incentives for take-home-tests raises the overall student's academic performance? We are in a context of multiple outcomes. So in order to draw general conclusions, in Table 2 we present findings of a summary index that aggregate information over the eight educational outcomes (*number of take-home-tests handed in, average grade at take-home-tests, grade of the midterm exam, grade at final exam, grade average in the midterm exams and homework of other simultaneous courses, average grade in the final exams of other simultaneous courses, accumulated grade average in the students career, total number of credits achieved in the semester*). To construct this summary index we follow the procedure used in Kling, Liebman and Katz (2007) and Dal Bó and Rossi (2011). This overall index is defined to be the equally weighted average of z-scores of its components, with the sign of each measure oriented so that more beneficial outcomes have higher scores[8]. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation.

[8] Summary Index = (percentage of take home tests + average grade at take home tests + grade at midterm exam + grade at final exam + average grade at take home tests & midterm exams of other simultaneous courses + average grade at the final exams of other simultaneous courses + accumulated average grade during the student's career + credits achieved in the semester)/8, all components built as z-scores.

[Insert Table 2]

Table 2 shows that the effect of Treatment 1 (Joint-liability Incentives), on the overall index that averages together all eight outcomes, is statistically significant and the size of this overall effect is around 0.45 standard deviations, in comparison with the control group[9] [10]. These results are similar when we control for the variables that are unbalanced due to attrition (*Interior as region of origin, number of friends in the class, number of totally unknown people in the class*)[11]. Given that grades at take-home-tests in the experimental courses may be too noisy (students may cheat due to the pressure exerted by the peer monitoring), we also build the index without the variable *average grade at take-home-tests* and the results are similar[12]. This positive average effect of the joint-liability mechanism is also present in other research areas like microfinance (Becchetti and Pisani, 2010; Banerjee and Duflo, 2010) where theory argues that this instrument gives poor borrowers strong incentives to monitoring each other and, thus, reduces moral hazard. One of the most important keys of success is considered to be the joint liability mechanism, that is, the bank provides small individual loans to a group of borrowers and enforces a contract in which an individual's default on repayment implies penalties for the other group-mates.

On the other hand, as Table 2 reports, Treatment 2 (Individual Incentives) has no significant effect on the students' performance in the course. This result is in line with previous literature that suggest that though grades may be theoretically valuable as an ability signal in the job market (Zubrickas, 2012), they are not effective motivators in college classes

[9] Table 2 considers 43 individuals due to, besides the six individuals who suffer attrition, two students did not take the final exam (they did not reach the minimum required grade of 4 at homework&midterm).
[10] The absolute magnitudes of the indices are in units akin to standardized test scores: the estimates shows where the mean of the treatment group is in the distribution of the control group in terms of standard deviation units.
[11] Results available from the authors upon request.
[12] Results available from the authors upon request.

at universities (Grant and Green, 2012), at least when they are designed as individual incentives.

The fact that Treatment 1 (Joint-liability Incentives) increases the index of overall performance may be the result of different patterns of effects over the individual outcomes that are included in the index. Thus, we investigate in Table 3 the effect of the treatments on each of the eight educational outcomes that are linked with the student's academic performance[13].

[Insert Table 3]

The first column of Table 3 reports the effects on the percentage of *take-home-tests handed in* by the students. Treatment 1 (Joint-liability Incentives) seems to impact positively on the homework done by the students, increasing the percentage of take-home-tests handed in by 18 percent, an increase of 30 percent relative to the control group. Treatment 2 (Individual Incentives) does not show any significant impact. In the second column, we observe the effect of the treatments on the average grade of the take-home-tests[14]. We standardize the results of the *average grade at take-home-tests* for each of the courses (Macroeconomics and Descriptive Economics) [15]. While Treatment 1 (Joint-liability Incentives) increases the standardized *average grade at take-home-tests* by .75, Treatment 2 (Individual Incentives) seems to have no effect. The third column shows us the impact of the treatments on the midterm examinations. We also standardize the results of the *grades in midterm exams* for each of the courses (Macroeconomics and Descriptive Economics). Those who

[13] The results are similar when we include no controls and when we control for the variables that are unbalanced due to attrition (*Interior as region of origin, number of friends in the class, number of totally unknown people in the class*). Results are available from the authors upon request.
[14] We do not include the prize of 20 percent in this average grade.
[15] The standardized grades are calculated by subtracting the course (Macroeconomics I or Descriptive Economics) mean and dividing by the course standard deviation. Average grades at take-home-test do not include the 20% premium.

received Treatment 1 (Joint-liability Incentives) outperform the control group by .7 in the standardized grades of midterm examinations. Once again, Treatment 2 (Individual Incentives) does not show any significant impact. In column four, we see that the estimates do not report any significant impact on the grade at final examination. At first sight, these findings could show that the positive impact of the group incentives is present only in the short run (higher percentage of take-home-test handed in with higher grades on average and higher grades at midterm exams) and fades out in the long run (there`s no improvement in the grade at the final exam among the students who receive the treatments). Moreover, one may state that this program of extra incentives may distort the quantity of time that students assign to the different courses of the semester. In other words, these incentives may divert the students efforts from other courses, condemning them to poorer results in the grades accomplished at other courses. In order to study this argument, we should find out the spillover effects of this program of extra incentives.

The fifth column of Table 3 reports the effects of the treatments on the average grade accomplished at midterm exams and homework of other simultaneous courses taken by the students in the same semester. Treatment 2 (Individual Incentives) does not show any significant impact, but Treatment 1 (Joint-liability Incentives) increases the average grade of midterm exams and homework of simultaneous courses by 1.16, an increase of nearly 20 percent relative to the control group.

Though in column sixth we observe that there`s no improvement in the average grade at the final exams in the other simultaneous courses among the students who receive the treatments, the seventh and eighth columns show positive spillover effects. The joint-liability incentives increase the accumulated grade average accomplished by the students in their

undergraduate life by nearly 12 percent in comparison to the control group. And Treatment 1 also increases the credits completed in the semester by 9, an increase of nearly 40 percent relative to the control group. Hence, Treatment 1 (Joint-liability incentives) increases the overall student's academic performance in the semester.

In sum, joint-liability incentives increase academic performance during the period of classes both in the experimental courses and in the other simultaneous courses of the semester. This positive effect dilutes during the period of exams though eventually the overall impact of group incentives on academic performance is positive. There are several explanations for this and we discussed them -after the follow up of the experiment- with a focus group formed by students who had participated in the experiment. This discussion was an enriching experience to evaluate different hypotheses. For instance, in terms of the model of Becker and Murphy (1988) - employed also by Charness and Gneezy (2009) in a field experiment about the formation of fitness habits - peer monitoring may increase human capital accumulation and develop habit formation[16]. This greater stock of human capital may have positive effects on the academic performance of all the courses in the semester but joint-liability incentives may not achieve to develop a strong habit of studying. Thus, the rate of disappearance of the human capital, the rate of preference for the present and the absence of strong habits of studying may explain the null effects of the treatment in the period of final exams –when the joint-liability incentives are absent[17].

An additional possible reason behind our results is a kind of peer effect. The relative better performance of students in midterm exams under peer-monitoring is a signal to the

[16] The motivation for the hypothesis "students will study more frequently after the incentives are removed as compared to before the incentives were introduced" is "habit formation".
[17] "Habits increase the marginal utility of engaging in an activity in the future. People seem to systematically underestimate the impact of their current actions on the utility of future action and to discount the future too much. As a result, people may underinvest in habit-forming activities" (Charness and Gneezy, 2009).

control group that they should study more for final exams and that they should obtain the class-notes of the treated students and study with them. Thus the control group may be catching up.

Given previous findings that show a positive effect of attendance on academic performance (i.e. Dobkin, Gil and Marion, 2010), one could argue that the positive effects of the joint liability scheme during the period of classes may be based on the possible higher rate of attendance of students under the pressure of peer monitoring. But, in this field experiment, attendance does not seem to be the cause of better performance since the students assigned to the joint liability treatment did not show a higher attendance rate[18].

Another potential explanation for our findings of no effects at the final exam may be that students just wish to accomplish a satisfactory performance in their overall academic semester, that is, in the four or five courses that they usually take per semester[19]. The instructor wishes to elicit high effort by them in his course. Under the pressure of a scheme of peer monitoring and joint liability, the students take on the startup cost –that may loom large at first sight- of coordinating to prepare take-home-tests with other classmates after school hours and sit down to study with them. Peer-monitoring moves some people past the "threshold" needed to really engage in learning, at least for some time. Once they have taken on this sunk cost, students devote time with their classmates not only to study for the experimental course but also to the other simultaneous courses of the semester due to they

[18] Results available from the authors upon request.

[19] Tommasi and Weinschelbaum (2007) suppose a certain principal-agent relationship where the principal (the instructor) offer a contract to the agents (students) to elicit high effort by them. The contract is designed as a scheme of peer monitoring. The agent accepts this contract but then unwinds part of these incentives through additional trades. Tommasi and Weinschelbaum refer to these outside trading opportunities as "insurance". The main function of these potential trades is to take risk away from the agents, hence playing an insurance role. In terms of our experiment, the students assigned to the treatment 1 (joint-liability incentives), are obliged, by means of peer monitoring, to increase their attention devoted to the course. But the students take not only the experimental course but also four or five courses per semester and they want to get a satisfactory overall performance; they are not interested in devoting a great attention to only one course.

seek satisfactory performance in their overall academic semester. Thus, treated students accomplished better academic performance at homework and midterm examinations during the period of classes. The positive academic experiences achieved during the period of classes may be a source of creating a sense of self-efficacy because they provide students authentic evidence that they have capability to succeed at the task (Dochy, Segers and van Dinther 2011), so at the time of the final exams, when the peer monitoring disappears, the students unwind the incentive to achieve better grades at final exams and rest on the higher grades achieved at homework and at midterms during the courses of the semester, adjusting downward the time devoted to study for the final exams. This downward adjustment is limited by the fact that each course at the university demands a minimum grade of six (in a one-twelve scale) in the final examination to approve the course. Hence, the overall academic performance at the semester increases since each course in the university is graded taking into account the grade at homework & midterm exam (50 percent) –that increases by peer monitoring- and the grade at the final exam (50 percent) –that is not affected by the treatment. In sum, the joint-liability incentive does not harm the performance in simultaneous courses, and really is effective to increase overall academic performance.

Exploiting the data available at the follow up survey, we are interested in measure if this Treatment 1 (Joint-liability Incentives) that achieved positive effects on students' global academic performance in the semester has spillover effects on students' subjective well-being.

[Insert Table 4]

As Table 4 shows, the group incentives impacts negatively on the satisfaction with classmates reported by the students[20]. This finding may capture that the students that receive the group incentives were assigned to groups of three students by randomization. That is, to win the prize of an extra grade of 20 percent demands that each one in the group of three fulfills the requirements (attendance to class, take-home-tests handed in, a minimum grade at the take-home-tests). If one of the three classmates of the group does not honor the requisites, all of them are condemned to lose the prize, no matter the individual effort done. In other words, many of these students are freshmen from different high schools of origin, they are not necessarily close friends and they are required to interact within a group. They may develop some reproaches to the other members of the group, for instance, each time one of them did not hand in the homework he makes the other members of the group lose their prize; or if one student of the group of three behaves as a free rider cheating the homework. But in these occasions, they may not have enough confidence to express their anger or frustration openly. Also the free riders may be resented because they are thought to be taking more than their fair premium or failing to shoulder any part of the cost of it. Thus, these hidden reproaches and resentments may manifest at the follow up survey. We think this is a novel result, and is not mentioned for example in Banerjee and Duflo (2010) as a cost of group liability schemes (one of such costs is imposing excessive risk-aversion on members, as in Banerjee, Besley and Guinnane, 1994).

[20] The results are similar when we control for the variables that are unbalanced due to attrition (*Interior as region of origin, number of friends in the class, number of totally unknown people in the class*). Results are available from the authors upon request.

As Table 5 reports, the students that received Treatment 1 does not seem to extend these reproaches to the evaluation of the instructor done by the students[21].

[Insert Table 5]

However, Treatment 2 (individual incentives) impacts negatively on the evaluation of the instructor of the course. Gneezy, Meier, and Rey-Biel (2011) provide a possible explanation stating that offering incentives for improved academic performance may signal that achieving a specific goal is difficult, that the task is not attractive, or that the agent is not well-suited for it, or that the principal does not trust the agent's intrinsic motivation. Also, the individual incentives design makes clear for the rest of the classmates if the student achieves the requirement. This increase in the signal may result in a lower personal image, and thus, contrary to what one could expect at the beginning of the experiment, the student may be unhappy with the instructor for the assignment to the treatment of individual incentives.

A usual concern in evaluations of programs by randomization is that results of the control group may be negatively affected by the effect of bad luck in the lottery on motivation. However, table 5 reports that the students who were assigned randomly to the control group do not show any significant difference in the evaluation of their instructor relative to the other groups.

Finally, we run a placebo test. We postulate that there is no plausible channel through which the program could affect the students' satisfaction with the neighborhood at which the university is located. Thus, we should see negligible effects on the outcome *satisfaction with the neighborhood of the university*.

[21] Given that the evaluation of professors is confidential information, we are not able to employ individual level data. Instead we have obtained aggregated data of the evaluation of each professor (Macroeconomics and Descriptive Economics) by treatment 1, treatment 2 and control group.

[Insert Table 6]

As we expected, we find no significant impact of group incentives treatment on the students' satisfaction with the neighborhood of the university[22]. Thus, we may infer that the previous findings (joint-liability incentives increase homework done and its average grade, grades at midterm exams, average grades in other courses, average grade in the student career and credits accomplished in the semester) are operating through the jointly liability mechanism and are not spurious correlations. This, together with the random assignment to treatment, leads us to believe in the causal interpretation of our previous findings.

IV. Conclusions and Discussion

Several conclusions emerge from this randomized field experiment. First, joint-liability incentives increase academic performance in the course by peer monitoring. Second, joint-liability incentives have positive spillover effects on the other simultaneous courses taken on by the treated students in the semester. Our results suggest that group incentives improve the overall index of academic performance in the semester. Both the direct effects and the spillover effects show a large percentage increase in comparison to the control group. The main drawback of these positive effects of the joint-liability incentives is the decrease in the rate of satisfaction reported by the treated students towards their classmates. Third, the program appears to be very cost-effective: we manage to design a successful mechanism to improve student´s academic achievements without giving monetary rewards. Fourth, individual incentives show no effect on academic performance, but seem to impact unfavorably on the evaluation of the instructor made by students. Fifth, while students under the joint-liability incentives outperformed the other students on homework and midterm

[22] The results are similar when we control for the variables that are unbalanced due to attrition (*Interior as region of origin, number of friends in the class, number of totally unknown people in the class*). Results are available from the authors upon request.

exams, there was no statistically significant improvement on the final exam. There are several possible explanations for this. The positive impact of joint-liability incentives diminish with time, or the control group may be caching up through peer or signaling effects, or the students may seek only a satisfactory performance in all the courses and not an special grade just in the experimental courses. Further research could help distinguish among these possibilities.

Another open question in a joint-liability scheme is the effect of class size and the effect of group size in the efficiency of peer monitoring. In this field experiment, class size is small and it makes easier to monitor the behavior of a classmate inducing her to do the homework properly, but in a larger class the cost of peer monitoring may be too high to be accomplished. For instance, we could imagine a class size of two hundred where it is very difficult just to know the name of each other. In addition, given all other conditions equal, a larger size of each joint-liability group certainly increases the cost of peer monitoring. In this field experiment, each joint-liability group is formed by only three students. But, what would happen if each joint-liability group is formed by nine students? A committed student could be discouraged by the greater probability that someone of the group may not fulfill the requirements to obtain the prize.

In the light of furthering our understanding, it is also important to study the long run impacts of the joint-liability incentives and the heterogeneity of effects among different students. What will happen if the additional incentive is reduced permanently? Will the effort be lower than it was before extrinsic incentives were offered? Negative long-run effects on students' joy of learning might be especially troublesome (Gneezy, Meier, and Rey-Biel, 2011). An interesting analogy is that in terms of incentives to sport exercising among undergraduate students, a strong decline –particularly on those who have already attend the

gym regularly- in exercising after removing the incentives is not completely rejected (Charness and Gneezy, 2009).

Finally, given the questionably efficacy of individual extrinsic incentives, educators may seek ways to make the learning experience more interesting, that is, if students develop intrinsic motivation for improve their knowledge and skills, they may become fully engaged with learning and devote more effort to this experience. Effort is shown to be important in improving the knowledge gained by students, and, by rewarding effort especially for certain students, it may motivate them to be better students (Swinton, 2010). This hypothesis requires more research.

The external validity of our conclusions is limited in principle to students similar to those that participate in this field experiment. Despite this selectivity, we should bear in mind that there is nothing atypical about these course characteristics, that are similar to firs year introductory courses in most universities. Certainly, it is unclear whether the conclusions of this research generalize to younger students. Hopefully subsequent investigations will clarify this. Designing systems to better accomplish the task of effectively motivating students represents a formidable challenge for researchers, policymakers, and educators. Our research fosters the literature on students incentives by suggesting that joint liability schemes should also be considered when designing such a system.

References

Angrist, J., Oreopoulos, Ph., & Williams, T. (2010). When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards. *NBER* Working Paper number 16643.

Angrist, J., Lang, D. & Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics,* 1(1), 136-163.

Banerjee, A., Besley, T. & Guinnane, T. (1994) Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test. *Quarterly Journal of Economics,* Vol. 109 (2), 491-515

Banerjee, A. & Duflo, E. (2010). Giving Credit Where It Is Due. *Journal of Economic Perspectives*, Vol. 24, No. 3, 61-79.

Becchetti , L. & Pisani, F. (2010). Microfinance, Subsidies and Local Externalities. *Small Business Economics* 34(3), 309-321.

Becker, G. S. & Murphy, K. M. (1988). A Theory of Rational Addiction. *Journal of Political Economy*, 96, 675-700.

Charness, G. & Gneezy, U. (2009). Incentives to Exercise. *Econometrica*, 77(3), 909-931.

Dobkin, C., Gil, R. & Marion, J. (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, 29, 566–575.

Dochy, F., Segers, M., & van Dinther, M. (2011). Factors Affecting Students' Self-efficacy in Higher Education. *Educational Research Review*, 6, 95-108.

Fryer, R. (2010). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *NBER* Working Paper number 15898. Cambridge, MA: National Bureau of Economic Research.

Dal Bó, E. & Rossi, M. (2011). Term Length and the Effort of Politicians. *Review of Economic Studies* 78(4), 1237-1263.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives,* 25(4), 1–21.

Grant, D. & Green, W. (2012). Grades as Incentives. *Empirical Economics,* forthcoming.

Kling, J., Liebman, J., & Katz, L. (2007). Experimental Analysis of Neighborhood Effects, *Econometrica* 75(1), 83-119.

Pozo, S., & Stull, Ch. A. (2006). Requiring a Math Skills Unit: Results of a Randomized Experiment, *American Economic Review* 96(2), 437-441.

Swinton, O. H. (2010). The effect of effort grading on learning. *Economics of Education Review*, 29, 1176–1182.

Tommasi, M. & Weinschelbaum, F. (2007). Principal--Agent Contracts under the Threat of Insurance. *Journal of Institutional and Theoretical Economics*. 163(3), 379--393.

Zubrickas, R. (2012). Optimal Grading (January 19, 2012). Available at SSRN: http://ssrn.com/abstract=989322 or http://dx.doi.org/10.2139/ssrn.989322

Table 1 - Pre-treatment characteristics by treatment assignment

	Treatment 1	Treatment 2	Control	Diff (Treat2-Treat1)	Diff (Control-Treat1)	Diff (Control-Treat2)
Age (in months)	238.904	233.757	237.605	-5.147	-1.298	3.848
				(5.165)	(5.614)	(4.576)
Male	.666	.785	.846	.119	.179	.060
				(.155)	(.154)	(.155)
Average grade	7.970	7.328	7.453	-.642	-.516	.125
				(.546)	(.521)	(.528)
Credits accomplished	53.333	35.642	48.423	-17.690	-4.910	12.780
				(15.549)	(18.157)	(15.015)
Bachelor in Economics	.541	.500	.538	-.041	-.003	.038
				(.172)	(.176)	(.199)
Work	.166	.214	.076	.047	-.089	-.137
				(.133)	(.120)	(.139)
Volunteering	.250	.214	.153	-.035	-.096	-.060
				(.146)	(.144)	(.155)
Interior	.250	.357	.307	.107	.057	-.049
				(.155)	(.156)	(.188)
High School 1	.291	.285	.230	-.005	-.060	-.054
				(.156)	(.156)	(.175)
High School 2	.166	.071	.076	-.095	-.089	.005
				(.115)	(.120)	(.104)
Hours of sports per week	3.812	5.178	4.423	1.366	.610	-.755
				(1.095)	(1.051)	(1.185)
Satisfaction with classmates	4.166	4.214	4.307	.047	.141	.093
				(.272)	(.260)	(.318)
Travel time to the university (in minutes)	27.708	27.142	22.692	-.565	-5.016	-4.450
				(4.667)	(4.649)	(3.786)
Group (1 = Macroeconomics;	1.500	1.500	1.461	.000	-.038	-.038
2 = Descriptive Economics)				(.172)	(.176)	(.199)
Study in group (in % of the time)	.280	.350	.411	.069	.131	.061
				(.078)	(.085)	(.094)
Friends (%)	.133	.184	.119	.051	-.013	-.064
				(.036)	(.036)	(.042)
Still unknown (%)	.557	.500	.588	-.056	.030	.087
				(.077)	(.084)	(.095)
Educational aspirations	3.875	4.000	3.461	.125	-.413	-.538
				(.320)	(.318)	(.386)
Observations	24	14	13			

Note: Standard errors in parentheses.

Table 2 – The effect of incentives on academic achievement

	Dependent variable: index of academic achievement		
	(1)	(2)	(3)
Treatment 1 Joint-liability	0.460 **	0.437 *	0.389+
	(0.202)	(0.218)	(0.234)
Treatment 2 Individualistic	0.189	0.165	0.205
	(0.225)	(0.241)	(0.295)
Controls:			
Gender	No	Yes	Yes
Age	No	Yes	Yes
Working	No	No	Yes
Time devoted to sports	No	No	Yes
Educational expectations	No	No	Yes
Observations	43	43	43

Robust standard errors in parentheses.

All models control by group where dummy=1 if student attends Macroeconomics group, and dummy=0 if attends Descriptive Economics group.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

+p value=0.106

Table 3 - The effect of incentives on academic achievement by outcome

	Effects on the course performance				Spillover effects			
	(1) Percentage of Take-home-tests handed in	(2) Average grade of take-home (standardized)	(3) Grade at midterm exam (standardized)	(4) Grade at final exam (standardized)	(5) Average grade in homework & midterm exams in other simultaneous courses	(6) Average grade at other simultaneous final exams	(7) Total average grade accumulated in the student career	(8) Credits accomplished in the semester
Mean of Control Group	***0.653***	***-0.524***	***-0.477***	***-0.006***	***6.252***	***7.056***	***6.646***	***23.346***
Treatment 1 Joint-liability	0.186**	0.635**	0.685*	-0.0249	1.153*	0.265	0.798*	9.229*
	(0.0761)	(0.263)	(0.371)	(0.460)	(0.678)	(0.602)	(0.465)	(5.353)
Treatment 2 Individualistic	0.0994	0.318	0.380	-0.102	0.235	0.162	-0.146	5.359
	(0.0895)	(0.310)	(0.470)	(0.495)	(0.919)	(0.713)	(0.519)	(6.521)
Controls: All models include gender, age, working status, time devoted to sports, educational expectations, group (dummy variable taking the value of 1 in Macroeconomics course).								
Observations	51	51	46	43	48	46	51	51

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 – The effects of incentives on satisfaction with classmates

	Dependent variable: Index of satisfaction with classmates		
	(1)	(2)	(3)
Mean of Control Group	***4.300***	***4.300***	***4.300***
Treatment 1 Joint-liability	-0.474	-0.513*	-0.502*
	(0.293)	(0.258)	(0.278)
Treatment 2 Individualistic	-0.0488	-0.0975	-0.0998
	(0.344)	(0.304)	(0.342)
Controls:			
Gender	No	Yes	Yes
Age	No	Yes	Yes
Working	No	No	Yes
Time devoted to sports	No	No	Yes
Educational Expectations	No	No	Yes
Observations	45	45	45

Robust standard errors in parentheses

All models control by group where dummy=1 if student attends Macroeconomics group, and dummy=0 if attends Descriptive Economics group.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 – The effects of incentives on the evaluation of the instructor made by students

	Treatment 1	Treatment 2	Control	Diff (Treat2 – Treat1)	Diff (Control – Treat 1)	Diff (Control –Treat 2)
Mean	0.239	-0.534	0.139	-0.773**	-0.100	0.673
				(0.347)	(0.339)	(0.439)
Observations	21	12	10			

Note: Standard errors in parenthesis

Table 6 – False experiment – Satisfaction with the neighborhood of the university

	(1)	(2)	(3)
Treatment 1 Joint-liability	-0.367	-0.317	-0.408
	(0.366)	(0.366)	(0.375)
Treatment 2 Individualistic	0.209	0.186	0.0930
	(0.380)	(0.400)	(0.380)
Controls:			
Gender	No	Yes	Yes
Age	No	Yes	Yes
Working	No	No	Yes
Time devoted to sports	No	No	Yes
Educational expectations	No	No	Yes
Observations	45	45	45

Robust standard errors in parentheses
All models control by group where dummy=1 if student attends Macroeconomics group, and dummy=0 if attends Descriptive Economics group.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$