

Joint-Liability vs. Individual Incentives in the Classroom

Lessons from a Field Experiment with Undergraduate Students

José María Cabrera*

Erasmus School of Economics & Universidad de Montevideo

Alejandro Cid**

Universidad de Montevideo

January 28th, 2014.

Abstract

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

JEL: I20, I23

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

*José María Cabrera jmcabrera@um.edu.uy; the corresponding author, Universidad de Montevideo, Prudencio de Pena 2544, Montevideo 11600, Uruguay, Tel. & Fax: +598 27074461. **Alejandro Cid acid@um.edu.uy.

We gratefully acknowledge the guidance of Martín Rossi. We also thank Ana Balsa, Marcelo Caffera, Juan Dubra, Nestor Gandelman, Eric Hanushek, Omar Licandro, Susanne Neckermann, Danilo Trupkin, Dinand Webbink and seminar participants at Universidad de Montevideo and XXVII Jornadas Anuales de Economía for valuable comments. All errors remain our own.

I. Introduction

Incentives for teachers have received considerable attention in previous literature, but less attention has been paid to encouraging students (Angrist et al 2014; Angrist et al 2009; Fryer 2011; Grant and Green, 2013) and the literature 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 throughout the history of educational systems, partly in response to demands for better information about undergraduate performance, but were never explicitly designed to motivate students (Grant and Green, 2013). We provide some ideas and results to design a system of incentives for students that does not require monetary rewards and makes them exert more effort in order to succeed in the course. .

One can think of a model in which the Principal (teacher) wants to exert effort from the Agents (students). For the design of such a contract (the grading rules of the course), we evaluate a novel design of student incentives, based on a joint-liability scheme. We borrowed the idea from the microfinance literature. One of the most important keys to success for the repayment rate in the microcredit experience 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. Theory argues that this instrument gives poor borrowers strong incentives to monitor each other and, thus, reduces moral hazard (Becchetti and Pisani, 2010; Banerjee and Duflo, 2010)

We designed a joint-liability contract that gave students strong incentives to monitor each other. We evaluated the impact of this contract using a randomized field experiment. The instructor designed groups of three students in the classroom and provided a premium to their homework's grade only if all three members of the group met some requirements. To

avoid self-virtuous group selection, participants were randomly assigned to each group. And, in order to distinguish the pure effect of peer monitoring from self-motivation, we also randomly assigned students to a group of individual incentives. Hence, applying randomization, we assigned students to the joint-liability treatment, to the individual-incentives treatment and to the control group.

We found that joint-liability incentives impacted positively on the grades achieved in homework and midterm exams both in experimental courses and in the other courses taken by students in the semester. Though the average positive effect seemed to disappear for final exams, the overall impact of joint-liability incentives on the academic achievements in the semester was still positive. On the other hand, the individual-incentive scheme had no effect. This result is in line with previous literature that provides no conclusive evidence about the effect of individual incentives on grades.

We also explored the channels that may be operating. For doing this, we entered the black box of the experiment with the help of a focus group of participants, and then tested several hypotheses for the results. We found that students did not care much about the grade (if it was above the minimum threshold that gives them the right to sit for the final exam), but they cared for the team members in the joint-liability design, so they put in more effort on their behalf. Although we cannot tease out empirically the different channels, with the help of the students participating in the focus group we discuss some other mechanisms: peer pressure, self-control problems, sunk costs of coordinating with class mates, etc.

We acknowledge that the sample size is small ($n=51$), but the behavioral changes created by the joint-liability incentives design were large enough to be statistically detected. The numerical findings were also confirmed when discussing the results with participants, in the focus group that was held when the experiment ended. We find the joint-liability design of incentives very interesting and promising. The significant effectiveness of any peer

monitoring developed by the joint-liability of group incentives provides novel implications for the design of grading policies in the classroom and for other social settings where incentives may be based in peer monitoring or joint liability. These ideas can be developed and tested in a broader class of issues, such as performance pay for teachers (where a joint-liability contract can be established, tying the earning of a bonus to the performance of other teacher's students, and thus favoring peer-monitoring between teachers), or to design a contracts between neighbors to reduce electric consumption in an apartment building, or between workers to increase output in a factory or office. In all these settings, peers (other teachers, neighbors, colleagues) can help the other members of the group meet the target (student grades, electric consumption, production), and thus joint-liability incentives could be better than individual ones if groups are small enough to reduce free-riding.

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

II. Program and experiment design

The experimental courses were core classes for freshmen at Universidad de Montevideo, a private university in Uruguay –a developing country in Latin America. The course composition was primarily undergraduate students majoring in Economics, Management and Accountancy. Undergraduate students at Universidad de Montevideo have to complete a number of credits in core courses in order to obtain their bachelor's degree¹. Two of these core courses are Macroeconomics I and Descriptive Economics, which students usually take during their first year at 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

¹ A different number of credits may correspond to each course. One credit corresponds to ten hours of class.

final exam (50%)². The minimum grade to pass the course is 6 in a scale from 1 to 12. Also, attendance to class is mandatory. Each course has sixty classes of fifty minutes each distributed throughout fifteen weeks and students may have up to fifteen absences. There is nothing atypical about the characteristics of these courses or the grading system in comparison with other courses offered at Universidad de Montevideo. We built a program that consists of giving incentives for take-home tests and attendance.

We wanted to test if incentives designed as a joint-liability scheme improved academic outcomes. We faced two major challenges to determine this causal effect. The first one was *self-virtuous group selection* (no one wanted to be grouped with lazy classmates to minimize the probability of losing the reward), which we overcame with the random assignment of participants to groups. A second challenge was that if faced with an incentive, an individual may put in more effort, *whether he is in a group or not*. If we only had the joint-liability treatment and a control group we would not have been able to distinguish between selfish motivations to get the prize and pure peer monitoring. So we built two different treatments in the classroom: individual and joint-liability incentives, and a control group. With this design we thought that we could identify the pure monitoring effect of peers.

Thus, we randomly distributed students in three groups³. In the *Joint-liability* group (Treatment group 1), the student was randomly assigned to a group of three and received a 20% increase in the grade of each take-home test if each student of his/her group fulfilled two conditions: he/she obtained a grade of at least 6 in the take-home test, and he/she had no absences during the week in which the take-home test had to be handed in.

² The frequency of take-home tests is nearly one every two weeks. Instructors determined this number of take-home tests looking for a sufficient number of occurrences that may form the habit of exercising.

³ The grading of these take-home tests is done by research assistants that do not know the distribution of students among the different treatments.

In the *Individual incentive* group (Treatment group 2), the student received a 20% increase in the grade of each take-home test if he/she obtained a grade of at least 6 in the take-home test, and he/she had no absences that week. These were the same requirements as for Treatment group 1, but they did not depend on the compliance of others⁴.

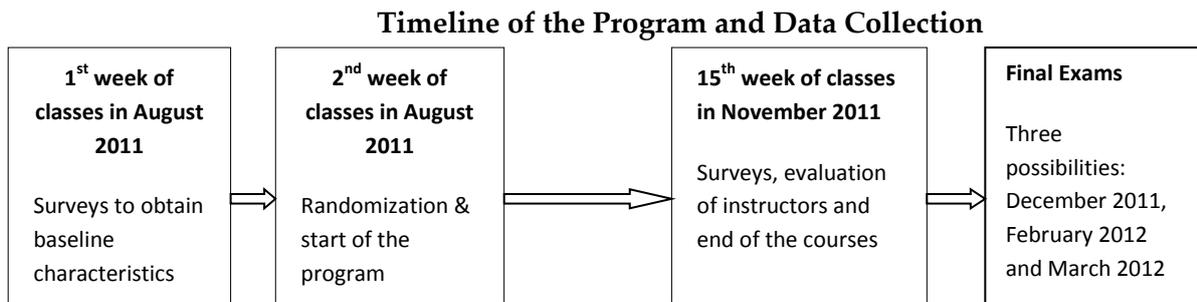
In the *Control* group, the student did not receive any incentives besides the general grading conditions of the course.

Take-home tests in this field experiment did not require team work, even for students in Treatment group 1: they could solve the problem sets alone, or with members of their team or with other students (from the control group or the other treatment). Each student was required to hand in his or her personal sheet with solutions at the beginning of the class and there was no problem if his or her solutions were identical to the ones of another classmate.

For the evaluation design we used randomized trials, with the approval of the ethical review board of the university. There were 51 different students in this field experiment: 26 in Macroeconomics I and 25 in Descriptive Economics. The selection process was as follows. In August 2011, all 51 applicants were asked to take part in a survey. In this baseline survey we collected data on a wide array of student characteristics such as age, gender, working hours, hours devoted to sports and volunteering, high school of origin, region of the country they come from, distance between their home in Montevideo and the university, academic expectations, and number of friends in the course. We also had administrative baseline data

⁴ The prize of 20% in the individual incentive group had a higher expected value than the prize in the joint-liability group. Suppose that all students are identical and have the same probability of not handing in a take-home test (a random negative shock to performance). Because the prize depends on the realization of the three random shocks in the joint-liability group, then the expected value is lower than in the individual case. The bigger individual prize goes against our findings (one of the results of the experiment is that the individual incentive did not work). The same prize of 20% was done for ease of exposition to students, and because when the experiment was designed we did not have information on the probability of not handing in a take-home test in order to calibrate the individual prize.

provided by the university such as average grade in previous courses, and number of credits already earned at the university. From this population, and given the restriction that the number of students in the *joint-liability* group had to be a multiple of three, 24 students were randomly assigned to Treatment group 1, 14 to Treatment group 2, and the remaining 13 candidates were assigned to the control group.



Once the random allocation was performed, the balancing condition was checked. If there were significant differences at the ten percent level in mean pre-treatment characteristics between the control and treatment groups, the random assignment procedure was repeated until we obtained an allocation that fulfilled the balancing condition.

[Insert Table 1]

Table 1 reports the balancing condition and shows that the three groups had similar characteristics. They were balanced in eighteen observable variables. By the random allocation design, the probability of receiving a treatment was orthogonal to students' characteristics. Therefore, including these characteristics in the regression model, while it may reduce standard errors, is not necessary for consistency⁵.

As usual in studies that follow students during the period of classes, some observations suffered attrition. In November 2011 two individuals from Treatment Group 1, one from Treatment Group 2, and three from the Control Group dropped out of the program. We had

⁵ Our findings do not change if we include controls in the estimates.

some outcomes from them during the courses and from follow-up administrative data, but we were not able to collect the complete data (grade in midterm exam, satisfaction with classmates or evaluation of the instructor) in these six cases due to different reasons (most students were freshmen and usually a certain number of them change to other degrees, some drop out of the course before the midterm exam, and some refuse to evaluate the instructor because the evaluation demands extra time out of class).

We compared the pre-treatment characteristics of the individuals that suffered attrition and those students who remained in the treatment/control groups. Since fifteen out of the eighteen variables remained balanced, the baseline data provided a measure of the similarity of these two groups. Only three variables were not balanced: students that are not from Montevideo, students with fewer friends, and students with more unknown people in the class tended to drop out more⁶.

III. Econometric model and results

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

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

where Y_i is one of the outcomes of interest for student i (*percentage of take-home tests handed in, average grade in take-home tests, grade in midterm exam, grade in final exam, average grade in midterm exams and homework of other simultaneous courses, average grade in final exams of other simultaneous courses, accumulated grade average in the student's career, total number of credits earned*

⁶ We included these variables in the regressions and our findings were not modified. They are available upon request.

in the semester), $T1_i$ is a dummy variable that takes the value of one if student i is assigned to Treatment group 1 and zero otherwise, $T2_i$ is a dummy variable that takes the value of one if student i is assigned to Treatment 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 student i belongs to the Macroeconomics course and zero otherwise, X_i is a matrix of student characteristics, and e_i is the error term. Given that no-compliers are not a problem, we can estimate this equation consistently with Ordinary Least Squares (OLS).

Prior research suggested that graded homework causes students to devote more effort than when they are assigned non-graded homework (Pozo and Stull, 2006). Does providing joint-liability and individual extra incentives for take-home tests raise the student's overall 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 aggregates information over the eight educational outcomes. To construct this summary index we followed the procedure used in Kling et al (2007) and Dal Bó and Rossi (2011). This overall index is defined as the equally weighted average of the z scores of its components, with the sign of each measure oriented so that more beneficial outcomes have higher scores⁷. Z scores are calculated by subtracting the control group mean and dividing by the control group standard deviation.

[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

⁷ Summary index = (percentage of take-home tests + average grade in take-home tests + grade in midterm exam + grade in final exam + average grade in take-home tests and midterm exams of other simultaneous courses + average grade in final exams of other simultaneous courses + accumulated average grade during the student's career + credits earned in the semester)/8, all components built as z scores.

overall effect is around 0.45 standard deviations, in comparison with the control group⁸ ⁹. 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*)¹⁰. Given that grades in take-home tests in the experimental courses may be too noisy (students may cheat due to the pressure exerted by peer monitoring), we also built the index without the variable *average grade in take-home tests* and the results were similar¹¹. 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).

On the other hand, as Table 2 shows, Treatment 2 (individual incentives) had no significant effect on the students' performance in the course. This result is in line with previous literature that suggests that though grades may theoretically be valuable as an ability signal in the job market (Zubrickas, 2013), they are not effective motivators in classes at universities (Grant and Green, 2013), 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, in Table 3 we investigate the effect of the treatments on each of the eight educational outcomes that are linked with the student's academic performance¹².

[Insert Table 3]

⁸ Table 2 considers 43 individuals due to the fact that, besides the six individuals who suffer attrition, two students did not take the final exam (they did not reach the required minimum grade of 4 in homework and midterm).

⁹ The absolute magnitudes of the indices are in units akin to standardized test scores: the estimates show where the mean of the treatment group is in the distribution of the control group, in terms of standard deviation units.

¹⁰ Results are available from the authors upon request.

¹¹ Results are available from the authors upon request.

¹² The results are similar when we include no controls and when we control for the variables that are unbalanced due to attrition. Results are available from the authors upon request.

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 have impacted positively on the homework done by students, increasing the percentage of take-home tests handed in by 18 percent, which represents an increase of 30 percent in comparison with the control group. Treatment 2 (individual incentives) did not show any significant impact. In the second column, we observe the effect of the treatments on the average grade of take-home tests¹³. We standardized the results of the *average grade in take-home tests* for each of the courses (Macroeconomics and Descriptive Economics)¹⁴. While Treatment 1 (joint-liability incentives) increased the standardized *average grade in take-home tests* by .64, Treatment 2 (individual incentives) seemed to have no effect. The third column shows us the impact of the treatments on midterm examinations. We also standardized the results of *grades in midterm exams* for each of the courses (Macroeconomics and Descriptive Economics). Those who received Treatment 1 (joint-liability incentives) outperformed the control group by nearly .7 in the standardized grades of midterm examinations. Once again, Treatment 2 (individual incentives) did not show a significant impact. In column four, we see that the estimates do not report a significant impact on the grade in the final examination. At first sight, these findings could show that the positive impact of group incentives is present only in the short run (a higher percentage of take-home tests handed in with higher grades on average and higher grades in midterm exams) and fades out in the long run (there is no improvement in the grade in the final exam among students who receive the treatments). Moreover, it may be stated that this program of extra incentives may distort the amount of time that students devote 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 achieved in other courses. In order

¹³ We do not include the 20% prize in this average grade.

¹⁴ The standardized grades are calculated by subtracting the course mean (Macroeconomics I or Descriptive Economics) and dividing by the course standard deviation.

to study this argument, we should find out the spillover effects of this extra incentives program.

The fifth column of Table 3 reports the effects of the treatments on the average grade achieved in midterm exams and homework of other simultaneous courses taken by the students in the same semester. Treatment 2 (individual incentives) did not show a significant impact, but Treatment 1 (joint-liability incentives) increased the average grade of midterm exams and homework of simultaneous courses by 1.15, an increase of nearly 20 percent in comparison with the control group.

Though in column six we observe that there was no improvement in the average grade in the final exams of the other simultaneous courses among the students who received the treatments, the seventh and eighth columns show positive spillover effects. The joint-liability incentives increased the accumulated grade average attained by the students in their undergraduate life by nearly .8, which means a 12 percent increase in comparison with the control group. And Treatment 1 also increased the credits earned in the semester by 9, an increase of nearly 40 percent with regard to the control group. Hence, Treatment 1 (joint-liability incentives) increased the student's overall academic performance in the semester.

In sum, joint-liability incentives increased academic performance during the period of classes both in the experimental courses and in the other simultaneous courses of the semester. This positive effect diluted during the period of exams though eventually the overall impact of group incentives on academic performance was 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 proposed by Becker and Murphy (1988) -applied 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 formation¹⁵. 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 succeed in developing strong study habits. Thus, the rate of disappearance of human capital, the rate of preference for the present, and the absence of strong study habits may explain the null effects of the treatment in the period of final exams – when joint-liability incentives are absent¹⁶.

An additional possible reason behind our results is a peer effect of some kind. The relative better performance of students in midterm exams under peer-monitoring is a sign to the control group that they should study more for final exams and that they should obtain the class-notes of 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 et al, 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 attendance rate of students under the pressure of peer monitoring. But, in this field experiment, attendance did not seem to be the cause of better performance since students assigned to the joint-liability treatment did not show a higher attendance rate¹⁷.

Another potential explanation for our findings of no effects on the final exam may be that students just wish to deliver a satisfactory performance in their overall academic semester,

¹⁵ 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”.

¹⁶ “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).

¹⁷ Results are available from the authors upon request.

that is, in the four or five courses that they usually take per semester¹⁸. The instructor wishes to elicit a high level of effort from them in his course. Under the pressure of a scheme of peer monitoring and joint liability, students accept the startup cost –which may seem large at first sight–of coordinating to prepare take-home tests with other classmates after school hours and sit down to study with them, and alleviate a problem of self-control, where the individual in fact cares but only with the help of peer pressure does he reach what he desires. Peer monitoring moves some people past the “threshold” needed to really engage in learning, at least for some time. Once they have taken in this sunk cost, students devote time with their classmates not only to study for the experimental course but also for the other simultaneous courses of the semester due to the fact that they seek to achieve a satisfactory performance in their overall academic semester. Thus, treated students delivered better academic performances in homework and midterm examinations during the period of classes. The positive academic experiences undergone during the period of classes may be a source for creating a sense of self-efficacy because they provide students with authentic evidence that they are capable of succeeding in the task (Dochy et al, 2011) so at the time of final exams when peer monitoring disappeared, students dismissed the incentive to obtain better grades in final exams and relied on the higher grades obtained in homework and in midterms during the course of the semester, adjusting the time devoted to studying for final exams downward. This downward adjustment is limited by the fact that each course at the university requires a minimum grade of six (in a one-through-twelve scale) in the final examination to pass the

¹⁸ Tommasi and Weinschelbaum (2007) suppose a certain principal-agent relationship where the principal (the instructor) offers a contract to the agents (students) to elicit a high level of effort from 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 Treatment 1 (joint-liability incentives) are obliged, by means of peer monitoring, to increase the effort they devote to the course. However, students take not only the experimental course but also four or five other courses per semester and they want to get a satisfactory overall performance; they are not interested in devoting a great deal of attention to only one course.

course. Hence, the overall academic performance during the semester improves owing to the fact that each course in the university is assessed taking into account the grade in homework and midterm exam (50 percent) –which increases by peer monitoring– and the grade in the final exam (50 percent) –which is not affected by the treatment. In sum, the joint-liability incentive does not harm the performance in simultaneous courses, and is really effective in increasing the student’s overall academic performance.

Exploiting the data available from the follow-up survey, we are interested in measuring if this Treatment 1 (joint-liability incentives) that achieved positive effects on students’ global academic performance in the semester has spillover effects on the students’ subjective well-being.

[Insert Table 4]

As Table 4 shows, group incentives impacted negatively on the satisfaction with classmates as reported by students¹⁹. This finding may reflect that the students who received joint-liability incentives were assigned to groups of three students by randomization. That is, to win the prize of an extra 20 percent, each one in the group of three needed to fulfill the requirements (attendance to class and minimum grade in take-home tests). If one of the three classmates of the group did not honor the requisites, all of them were doomed to lose the prize, despite the individual effort made. In other words, many of these students are freshmen from different high schools of origin, and they are not necessarily close friends, but they are required to interact within a group. At times they may develop some reproaches towards the other members of the group. For instance, every time one of them did not hand in the homework, he/she made the other members of the group lose their prize; or if one student

¹⁹ The results are similar when we control for the variables that are unbalanced due to attrition. Results are available from the authors upon request.

from the group of three behaved as a free rider cheating in homework. But in these occasions, they may not have had enough confidence to express their anger or frustration openly. Also free riders may be resented because they are thought to be taking more than their fair premium or failing to shoulder any part of its cost. Thus, these hidden reproaches and resentments may manifest in the follow up survey. We think this is a novel result, since it is not mentioned in the literature as a cost of group liability schemes (Banerjee and Duflo, 2010; Banerjee et al, 1994).

As Table 5 reports, students who received Treatment 1 did not seem to extend these reproaches to the evaluation of the instructor²⁰.

[Insert Table 5]

However, Treatment 2 (individual incentives) impacted negatively on the evaluation of the instructor of the course. Gneezy et al (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, 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 it clear for the rest of the classmates whether the student meets the requirement or not. This increase in the signal may result in a lower personal image –individualistic person- and thus, contrary to what one could expect at the beginning of the experiment, the student may be unhappy with the instructor for being assigned to the individual-incentives treatment.

A usual concern in the evaluation of programs by randomization is that results from the control group may be negatively affected by the effect of bad luck in the lottery on motivation.

²⁰ Given that the evaluation of professors is confidential information, we are not able to use individual level data. Instead we have obtained aggregated data from the evaluation of each professor (Macroeconomics and Descriptive Economics) according to Treatment 1, Treatment 2 and Control group.

However, in Table 5 we report that students who were assigned randomly to the control group did not show a significant difference in the evaluation of their instructor in comparison with the other groups.

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

[Insert Table 6]

As expected, we found no significant impact of the joint-liability scheme of incentives on the students' satisfaction with the university's neighborhood²¹. Thus, we may infer that the previous findings (joint-liability incentives increase homework done and their average grade, grades in midterm exams, average grades in other courses, average grade in the student career and credits earned in the semester) are operating through the joint-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 as a result of peer monitoring. Second, joint-liability incentives have positive spillover effects on the other simultaneous courses attended by treated students in the semester. Our results suggest that joint-liability incentives improve the overall index of academic performance in the semester. Both the direct effects and the spillover effects show an important percentage increase in comparison with the control

²¹ The results are similar when we control for the variables that are unbalanced due to attrition. Results are available from the authors upon request.

group. The main drawback of these positive effects of joint-liability incentives is the decrease in the rate of satisfaction towards their classmates reported by treated students.

Third, the program appears to be very cost-effective: we managed 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 negatively the evaluation of the instructor done by students. Fifth, while students with joint-liability incentives outperformed the other students on homework and midterm exams, there was no statistically significant improvement on the final exam. There are several potential explanations for this, such as the possibility that the positive impact of joint-liability incentives may diminish over time; or that the control group may be catching up the treated through peer or signaling effects; or that students may seek only a satisfactory performance in all the courses and not a special grade just in the experimental courses. Further research could help make distinctions between 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 was small and it made it easier to monitor the behavior of a classmate inducing him/her to do homework properly, but in a larger class the cost of peer monitoring may be too high to be met. For instance, we could imagine a class size of two hundred where it is very difficult just to know each other's names²². In addition, and considering all other conditions remain the same, joint-liability groups of a larger size would certainly increase the cost of peer monitoring. In this field experiment, each joint-liability group was formed by only three students. But, what would happen if each joint-liability group was formed by nine students? A committed student

²² In some large (public) organizations, incentives are based on the performance of an entire department with hundreds of employees, or even the bonus is earned if all the departments of the organization meet some target. These kinds of contracts designed as joint-liability schemes, are unlikely to work due to free-riding and the huge cost of monitoring members of large groups.

could be discouraged by the greater probability that someone in 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 joint-liability incentives and the heterogeneity of effects on 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 et al, 2011). An interesting analogy can be made with incentives for sport exercising among undergraduate students, because the hypothesis of a strong decline in exercising after removing the incentives –particularly among those who already attended the gym regularly– is not completely rejected (Charness and Gneezy, 2009).

Another line of research is to try to separate the peer monitoring effect from the cooperation or team work effect. The joint-liability scheme that we have used to favor peer monitoring also may increase cooperation between students. For instance, in the focus group we discovered that committed student encouraged the lazy members of the joint-liability group, who had self-control problems to sit down to study, to study together so as not to lose the bonus. The joint-liability scheme thus reduced the cost to cooperate as a team once the students meet to exchange the results of the take-home tests. Although working together in groups was not a requirement to earn the bonus and despite of the precautions we took (the members of each joint-liability group were selected randomly, thus they were not necessarily friends nor prone to work together), team work is still one of the possible channels behind our results. To disentangle these effects, in a further research we aim to randomly assign *different* take-home tests within the members of some joint-liability groups, while in other groups the take-home tests will be the same for all the members. So in the former case, cooperation will

be more costly than in the latter, and hence, the chance of isolating the peer monitoring effect will be greater.

Finally, given the questionable efficacy of individual extrinsic incentives, educators may seek ways to make the learning experience more interesting; that is, if students develop an intrinsic motivation to improve their knowledge and skills, they may become fully engaged with learning and devote more effort to this experience. Effort is known to be important in improving the knowledge gained by students and, by rewarding the efforts of certain students in particular, 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 participated in this field experiment. Despite this selectivity, we should bear in mind that there is nothing atypical about these course characteristics, which are similar to first year introductory courses in most universities. Certainly, it is unclear whether the conclusions of this research may be generalized 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 furthers the literature on student incentives by suggesting that joint liability schemes should also be considered when designing such a system.

References

- Angrist, J., Oreopoulos, Ph., and Williams, T. (2014).** When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards. *Journal of Human Resources*, forthcoming.
- Angrist, J., Lang, D. and 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. and Guinnane, T. (1994).** Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test. *Quarterly Journal of Economics*, 109(2), 491-515.
- Banerjee, A. and Duflo, E. (2010).** Giving Credit Where It Is Due. *Journal of Economic Perspectives*, 24(3), 61-79.
- Becchetti, L. and Pisani, F. (2010).** Microfinance, Subsidies and Local Externalities. *Small Business Economics* 34(3), 309-321.
- Becker, G. S. and Murphy, K. M. (1988).** A Theory of Rational Addiction. *Journal of Political Economy*, 96, 675-700.
- Charness, G. and Gneezy, U. (2009).** Incentives to Exercise. *Econometrica*, 77(3), 909-931.
- Dobkin, C., Gil, R. and 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., and van Dinther, M. (2011).** Factors Affecting Students' Self-efficacy in Higher Education. *Educational Research Review*, 6, 95-108.
- Fryer, R. (2011).** Financial Incentives and Student Achievement: Evidence from Randomized Trials. *Quarterly Journal of Economics*, 126, 1755-1798.
- Dal Bó, E. and Rossi, M. (2011).** Term Length and the Effort of Politicians. *Review of Economic Studies*, 78(4), 1237-1263.
- Gneezy, U., Meier, S., and Rey-Biel, P.(2011).** When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4), 1-21.
- Grant, D. and Green, W. (2013).** Grades as Incentives. *Empirical Economics*, 44:1563-1592.
- Kling, J., Liebman, J., and Katz, L. (2007).** Experimental Analysis of Neighborhood Effects, *Econometrica*, 75(1), 83-119.
- Pozo, S., and 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. and Weinschelbaum, F. (2007). Principal-Agent Contracts under the Threat of Insurance. *Journal of Institutional and Theoretical Economics*. 163(3), 379-393.

Zubrickas, R. (2013). Optimal Grading. Available at SSRN-id2222645

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 (5.165)	-1.298 (5.614)	3.848 (4.576)
Male	.666	.785	.846	.119 (.155)	.179 (.154)	.060 (.155)
Average grade	7.970	7.328	7.453	-.642 (.546)	-.516 (.521)	.125 (.528)
Credits earned	53.333	35.642	48.423	-17.690 (15.549)	-4.910 (18.157)	12.780 (15.015)
Bachelor in Economics	.541	.500	.538	-.041 (.172)	-.003 (.176)	.038 (.199)
Work	.166	.214	.076	.047 (.133)	-.089 (.120)	-.137 (.139)
Volunteering	.250	.214	.153	-.035 (.146)	-.096 (.144)	-.060 (.155)
Interior	.250	.357	.307	.107 (.155)	.057 (.156)	-.049 (.188)
High School 1	.291	.285	.230	-.005 (.156)	-.060 (.156)	-.054 (.175)
High School 2	.166	.071	.076	-.095 (.115)	-.089 (.120)	.005 (.104)
Hours of sports per week	3.812	5.178	4.423	1.366 (1.095)	.610 (1.051)	-.755 (1.185)
Satisfaction with classmates	4.166	4.214	4.307	.047 (.272)	.141 (.260)	.093 (.318)
Travel time to university (in minutes)	27.708	27.142	22.692	-.565 (4.667)	-5.016 (4.649)	-4.450 (3.786)
Group (1 = Macroeconomics; 2 = Descriptive Economics)	1.500	1.500	1.461	.000 (.172)	-.038 (.176)	-.038 (.199)
Study in group (in % of the time)	.280	.350	.411	.069 (.078)	.131 (.085)	.061 (.094)

Friends in the classroom (%)	.133	.184	.119	.051 (.036)	-.013 (.036)	-.064 (.042)
Still unknown (%)	.557	.500	.588	-.056 (.077)	.030 (.084)	.087 (.095)
Educational aspirations	3.875	4.000	3.461	.125 (.320)	-.413 (.318)	-.538 (.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 ** (0.202)	0.437 * (0.218)	0.389+ (0.234)
Treatment 2 Individual	0.189 (0.225)	0.165 (0.241)	0.205 (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 the course taken by students (Macroeconomics or Descriptive Economics).

+Significant at the 10.7% level; *Significant at the 10% level; **Significant at the 5% level;

***Significant at the 1% level

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

	Effects on the course performance				Spillover effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Percentage of take-home tests handed in	Average grade of take-home tests (standardized)	Grade in midterm exam (standardized)	Grade in final exam (standardized)	Average grade in homework & midterm exams in other simultaneous courses	Average grade in other simultaneous final exams	Total average grade accumulated in the student's career	Credits earned in the semester
Treatment 1 Joint-liability	0.186** (0.0761)	0.635** (0.263)	0.685* (0.371)	-0.0249 (0.460)	1.153* (0.678)	0.265 (0.602)	0.798* (0.465)	9.229* (5.353)
Treatment 2 Individual	0.0994 (0.0895)	0.318 (0.310)	0.380 (0.470)	-0.102 (0.495)	0.235 (0.919)	0.162 (0.713)	-0.146 (0.519)	5.359 (6.521)
Controls: All models include gender, age, working status, time devoted to sports, educational expectations and course.								
Observations	51	51	46	43	48	46	51	51

Robust standard errors in parentheses.

*Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level

Table 4 – The effects of incentives on satisfaction with classmates

	Dependent variable: Index of satisfaction with classmates		
	(1)	(2)	(3)
Treatment 1 Joint-liability	-0.474 (0.293)	-0.513* (0.258)	-0.502* (0.278)
Treatment 2 Individual	-0.0488 (0.344)	-0.0975 (0.304)	-0.0998 (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 the course taken by students (Macroeconomics or Descriptive Economics).

*Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level

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

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

Note: Standard errors in parentheses.

Table 6–False experiment – Satisfaction with the university’s neighborhood

	(1)	(2)	(3)
Treatment 1 Joint-liability	-0.367 (0.366)	-0.317 (0.366)	-0.408 (0.375)
Treatment 2 Individual	0.209 (0.380)	0.186 (0.400)	0.0930 (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 the course taken by students (Macroeconomics or Descriptive Economics).

*Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level