GENDER IN A CONTEXT OF JOINT-LIABILITY INCENTIVES IN THE CLASSROOM

Marianne Bernatzky, José María Cabrera and Alejandro Cid

Universidad de Montevideo

Author Note

Marianne Bernatzky, Department of Economics and Center of Applied Research on Economics, Universidad de Montevideo; José María Cabrera, Department of Economics and Center of Applied Research on Economics, Universidad de Montevideo; Alejandro Cid, Department of Economics and Center of Applied Research on Economics, Universidad de Montevideo.

Correspondence concerning this article should be addressed to Alejandro Cid, Department of Economics, Universidad de Montevideo, Prudencio de Pena 2544, Montevideo 11600, Uruguay.

E-mail: acid@um.edu.uy

Abstract

As recently documented in health literature, incentives to exercise are effective in developing healthy habits. We evaluate this finding in a novel context (the classroom) applying an incentives scheme from microfinance (joint-liability incentives). We set up groups of three students and provided a premium to their homework's grade if all the members of the group met certain requirements. We investigated how the frequency of take home tests affects the students' academic outcomes, and explore if male and female students respond differently to a joint-liability incentives scheme. We find that this treatment improves the cumulative grade point average of male students, but not that of females. This finding expands on the literature on joint-liability and gender behavior in a novel context.

Keywords: gender; field experiment; classroom incentives; evaluation; joint-liability incentives

* mbernatzky@correo.um.edu.uy; jmcabrera@um.edu.uy; acid@um.edu.uy

Incentives to develop habits

Recent evidence from health literature indicates that financial incentives are effective in developing healthy habits and improving health indicators. Charness and Gneezy (2010) find that it may be possible to encourage exercise by providing a monetary compensation for attending a gym a given number of times during the course of one month. Their findings support the 'habit formation' hypothesis, that suggests that one's utility from consumption depends on one's past consumption. If one's current consumption of the good increases, one's future consumption of the good increases as well because the present consumption of the good raises the marginal utility of future consumption (Becker and Murphy, 1988). They contrast this hypothesis to the 'crowding-out effect'. Assuming that participants are initially intrinsically motivated to exercise, the extrinsic intervention could be counterproductive and destroy their initial motives to exercise. If at first they exercised because they felt it was good for their health, once the incentives are introduced, they might feel they do it just for the money. Hence, the intrinsic motives are destroyed. They conclude that habits increase the marginal utility from exercising and therefore participants of the experiment engage in more physical activities in the future.

The same relevant question of whether incentives are effective in developing good habits arises in the field of education. Focusing on previous literature on education, we find that monetary rewards do not usually motivate students (Angrist, Oreopoulos, and Williams, 2010; Angrist, Lang, and Oreopoulous 2009; Fryer 2010) and grades do not explicitly encourage them (Grant and Green, 2012). It may be the case that the existing incentives are not significant or effective enough to improve the academic performance of students. It might also be possible that students are aware of the benefits of studying, but are not capable of improving their performance without external support.

Nevertheless, there is recent evidence supporting joint-liability schemes as a way of providing incentives on students. As the teacher wants to exert effort from his students, he designs a contract by which they have incentives to monitor each other. Cabrera and Cid (2013) find that a joint-liability scheme- in comparison to individual incentives and control groups - impacts positively on grades of take-home tests and midterm exams, but not on the finals. In other words, joint-liability incentives may not succeed in developing strong study habits as the effects dilute in the long run.

Considering the benefits of exercising more on health and the positive effects on educational performance of a joint-liability framework, we design an experiment to explore whether the frequency in which the take-home tests are assigned may cause differences in students' performance under a joint-liability scheme.

Thus, in each classroom, the instructor set up groups of three students and provided a premium to their take-home tests' grade if all the members of the group met certain requirements. To test how the frequency of these take-home tests could affect the academic performance of students, two strategies were implemented: eight take-home tests (low frequency) and sixteen take-home tests (high frequency). To prevent students from self-virtuous group selection, participants were randomly assigned to each group.

Following Becker and Murphy (1988) and keeping in mind Charness and Gneezy (2009) findings in fitness intervention, we expect students with high frequency take-home tests to improve their academic performance in comparison to those with low frequency take-home tests, as a result of being more frequently exposed to exercises. We not only analyzing students´ performance in the intervention courses, but also the spillover effects, i.e. the effect on the overall academic outcomes.

We find that there are no differences in the educational outcomes of the students between the high and low frequency groups. However, the sign of the coefficients are plausible, in line with Becker and Murphy (1988). Possible reasons for these results are the small sample size (which was exacerbated by the attrition experienced in the evaluation), that the difference in the frequency between groups may not be enough to generate differences in effort, and the spillover effects between groups due to the exchange of solutions of take-home tests´ between students. These are lessons upon which to improve for future field experiments in similar contexts. In addition, given the small sample size, we address the concerns over statistical inference implementing permutation tests and the results remain unchanged.

In the second part of the paper, we explore the impact of a joint-liability framework on the gender gap on students´ cumulative grade point average. For this section we pool data from two experiments to increase power. Our results show that the incentives designed as a joint-liability scheme improve students' performance. However, this overall positive effect masks gender disparities, since it is concentrated among male students who improved their academic outcomes significantly. There is no evidence of an effect of the joint-liability incentives on female students. This result is in line with Duflo et al. (2013) and Crépon et al. (2011) who find no significant joint-liability effects on women empowerment.

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 results, section IV the discussion, and V the conclusion.

Program and Experiment Design

The experimental courses at Universidad de Montevideo (a private university in Uruguay) were taken primarily by freshmen students majoring in Economics, Management and Accountancy. These courses were Macroeconomics I and Descriptive Economics, both core courses at the University. These were structured in the same way in the 2012 academic year: a midterm exam (35% of the final grade), take-home tests (15%) and a final exam (50%). Each course consists of sixty classes of fifty minutes, each distributed throughout fifteen weeks, and students are allowed to have up to 15 absences. Both Macroeconomics I and Descriptive Economics share similar characteristics in the grading system with other courses at the University.

In Table 1 we define the variables used in the paper and present a set of descriptive statistics. Our sample consists of 48 students over 18 years of age with a mean average grade of 7.5; two thirds are from the inland of the country, nearly 42% are female and approximately 30% of the students come from two private high schools in Montevideo. Nearly 10% of the students in the classroom have a job, 19% do volunteer work and, on average, the students in the intervention practice sports 5 hours a week. Also, students are equally distributed between the Macroeconomics and Descriptive Economics courses. With respect to the student's social behavior, we find that they devoted 33% of their time to studying in groups. When students were asked about the number of classmates they considered friends, we find that, on average, 13% of them are so. Similarly, the average percentage of unknown classmates is 60%.

The aim of the intervention is to test whether high frequency take-home tests improve academic performance or not under a joint-liability scheme. With the approval of the ethical review board of the university, and in order to evaluate the intervention, we designed a randomized trial. Students were randomly assigned to two groups. Using this evaluation design we avoided self-virtuous group selection that could have grouped lazy students in the low frequency or control group. In the high frequency take-home tests group (Treatment Group), the student was randomly assigned to a group of three and received a 20% increase in the grade if each student in the group fulfilled the following conditions: obtained a grade of at least 6 in the take-home test and had no absences during the week in which the take-home test had to be handed in. This group was assigned 16 take-home tests during the 15 weeks of class.

In the low frequency take-home tests group (Control Group), the students were randomly assigned to a group of three and had to meet the same criteria in order to get the bonus of 20% increase in the grade. The only difference was that these students had 8 take-home tests during the course of the 15 weeks of class.

For both treatment and control group, take-home tests did not require team work, but they were allowed to prepare the take-homes with another classmate. At the beginning of the class each student was required to hand in the solutions in a personal sheet. It is important to note that the content of the 16 take-home tests is exactly the same as the content of the 8 take-home tests; the difference between the two groups relies in the distribution of tasks in the 15 weeks of class. So, the treatment is a variation only in the *frequency* of the exercises, not in the *total amount* of exercises each student has to complete during the course.

There were 48 students in this field experiment: 24 in Macroeconomics and 24 in Descriptive Economics. As Figure 1 shows, in August 2012, all 48 applicants were asked to complete a survey. Thus, we collected baseline data on a wide array of students characteristics such as age, gender, working hours, hours devoted to sports and volunteering, high school of origin, region of the country they came from, commute time to university, academic expectations and number of friends in the classroom. Then, 24 students were randomly assigned to the high frequency take-home tests group and 24 students to the low frequency take-home tests group.

Table 2 shows that the randomization appears to have been successful, since the three groups were balanced in eighteen observable variables.

We should mention that attrition did occur. It is not rare that students drop out from some courses during the semester due to different reasons (e.g. freshmen students usually change to other degrees and some students drop out before taking the midterm exam). In November 2012, 6 students dropped out of the program (2 students from the Treatment group and 4 students from the Control group). We have some outcomes and follow-up administrative data for those who suffered attrition, but we could not collect the information on all the outcomes of interest for the whole sample (e.g. grade in midterm exam). Therefore, taking this into account, we compared pre-treatment characteristics of the individuals that suffered attrition and the students that remained in the treatment/control groups, resulting in all the variables remaining balanced[1].

In this field experiment we work with a small sample of students. This could lead to difficulty in identifying significant impacts and to raise concerns over statistical inference in small samples. In order to address the last concern we implement permutation tests. The advantage of this method is that it is valid even with small samples and does not rely on the distributional characteristics of the data (Rossi, 2014; Bloom, Eifert, Mahajan, McKenzie, Roberts, 2013). We perform 10,000 permutation tests and calculate the t-statistics for every outcome. Once this is estimated, we obtain the 2.5% ($t_{0.025}$) and 97.5% ($t_{0.975}$) percentiles of the t-test distribution. In the case that we find that the t-values are greater than $t_{0.975}$ or smaller than $t_{0.025}$, it can be stated that the difference is significant at the 5% level and therefore the treatment effect is significant. This approach does not alter the results found in our research.[2]

Method

High Frequency vs. Low Frequency Take- Home Tests

[1] Results are available from the authors upon request.
[2] Results are available from the authors upon request.

The aim of this study is to estimate the causal effect of high frequency take-home tests on student's achievements. Formally, we estimate the following equation:

$$Y_i = a + bT_i + dMacro_i + X_i{}'f + e_i$$

where Y_i is one of the outcomes of interest for student i (grade in midterm exam, grade in final exam, average grade in take-home tests[3], average grade in midterm exams and homework of other simultaneous courses, average grade in final exams of other simultaneous courses, the student's cumulative grade point average)[4], T_i is the parameter of interest: a dummy variable that takes the value of one if student i is assigned to High Frequency Treatment (16 take-home tests) and zero otherwise, $Macro_i$ 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. Because of the random assignment to the treatment and the inexistence of non-compliers, we estimate the equation using Ordinary Least Squares (OLS).

The question is whether high frequency take-home tests in a joint liability framework raises the student's academic performance. In table 3 we investigate the effect of the treatment on the educational outcomes (*grade in midterm exam, grade in final exam and average grade in take-home tests*) in comparison to the control group. There are no significant differences between the two groups. However, there is an effect of the

---

[3] These three grades are standardized. Standardized grades are calculated by subtracting the course mean (Descriptive Economics or Microeconomics I) and dividing by the course standard deviation. We do not include the 20% prize in the average grade in take-home tests.
[4] In Panel B of Table 1 we present a description of the outcome variables at the follow-up.

treatment on *midterm exams*, but only if the differences are taken at the fifteen percent level.

In addition to this, it could be argued that the high frequency of the take-home tests on the experimental courses may have worsened the educational outcomes on other courses at the university (treated students diverted effort from other subjects in order to earn the 20% bonus). Thus, we study the following outcomes: *average grade in midterm exams and homework of other simultaneous courses, average grade in final exams of other simultaneous courses, and the student's cumulative grade point average*. We find that there are no differences between the control group and the treatment group.

Though we do not find significant effects, the coefficient of the treatment dummy (16 take-home tests) has the expected sign in the regressions. Thus, the treatment (high frequency take-home tests) seems to be positively associated with an improved average grade in take-home tests and with higher standardized grades in midterm exams and standardized grades in the final exams. Taking into account the spillover effects on the overall academic performance, though the effects are not significant, the signs of the coefficients of the treatment dummy are consistent with the initial hypothesis that a higher frequency in take-home tests impacts positively on academic achievement not only on the treatment courses but also on other simultaneous courses.

One possible explanation for the lack of significance is that there were only 48 students in the experiment. The results were also weakened due to attrition in the intervention: given our sample size, six students represent a significant attrition rate.

Another reason could be that, although students in the low frequency group had to hand in their tasks every two weeks, they could get a head start on their tasks by studying with those in the treatment group (positive spillover effects), leading to no differences in habit formation from arising. In this sense, it could be argued that spillover effects did occur. Also, it could be the case that the difference in the frequency of take-home tests is too small to be able to see any differences in effort among students; or that the duration of the intervention (15 weeks) is too short of a time to see changes in habit formation. Finally, we do not rule out the possibility that the only mechanism working is the joint-liability incentive and that the frequency is not a relevant factor. We properly discuss each of these arguments in the last section of this paper.

Gender & Academic Outcomes

Gender differences have been widely documented. Men have more attention and behavioral difficulties, lower levels of inhibitory control and perceptual sensitivity and are more likely to be diagnosed with attention deficit hyperactivity disorder (ADHD) (Bertrand and Pan, 2013; Ruigrok, et al., 2014). Whether this gender gap in non-cognitive skills is determined by biological differences or social influences is unclear. Biological differences are associated to differences between male and female brain structures in areas related to mood, emotions and emotional regulation. Social influences may be related to home and school environments (Bertrand and Pan, 2013). We are particularly interested in gender because it is highly correlated to non-cognitive skills, which in turn might be a determinant of educational achievements.

We focus on the effect that a joint-liability framework may have on the cumulative grade point average considering the interaction term with gender. Cabrera and Cid (2013) find that joint-liability incentives are effective in improving academic performance. However, whether these incentives are beneficial to both female and male students is uncertain. Previous literature in microfinance finds no significant joint-liability effects on women empowerment (Duflo et al., 2013; Crépon et al., 2011). Therefore, we expect no significant changes in the cumulative grade point average of female students.

For this section we make use of an experiment conducted the year before the current experiment. We take advantage of the facts that we designed the two experiments with the same joint-liability framework, that they were implemented in the same courses and with similar populations, that they share the same baseline survey, and that we have a homogeneous outcome for both years: average grades. The main difference between the two experiments is that in Cabrera and Cid (2013) the focus was to evaluate the effects of joint-liability incentives on educational outcomes in comparison to individual incentives and a control group, and in the new experiment presented in the first part of the paper, the interest was on the effect of high vs low frequency tasks. It is important to notice that we designed both experiments in such a way that they share the same joint liability framework, so we can pool the data from both years and have a joint-liability treatment randomly assigned to students. Table 4 shows the balance in 17 pre-treatment variables for this experiment (that pools data from two years) controlling for a year dummy since randomization was performed separately for each year (and not with the pooled data). There are no discernible differences between students treated with joint-liability incentives and the control group. The only disparity between students is that some of them were treated in the first year and the rest in the second year.

After establishing the validity of the research design with the random assignment of the treatment, we will present baseline descriptive statistics by gender. In table 5 we describe female and male students at baseline. Male and female students are not different in age or region of origin of the country, they have similar educational aspirations, devote a similar share of their time to studying in groups, commuting to university and volunteering in social activities. They are also similar in terms of the percentage of them that have a job; and the satisfaction with classmates is much the same. However, male students tend to spend more time practicing sports, have more friends in the classroom and less classmates they do not know. Female students have a higher cumulative grade point average – the difference is almost one point: 8.4 vs. 7.5- and a higher percentage of them are majoring in economics. We also find that a higher percentage of male students are taking the macroeconomics course. Moreover, the number of women taking part in the experiment is higher in the second year. When we control for year, male and female, the same differences at baseline arise.

The outcome of interest is the cumulative grade point average for the student's career, which was homogenously measured in both years. This variable provides an overall picture of the students' performance and we exploit the fact that it was measured before and after the joint-liability treatment.

Taking this into account, we report the effect of being in a joint-liability framework on the difference experienced in the cumulative grade point average. After that, we estimate the effect of being in a joint-liability framework considering an interaction term with gender.

In table 6 we present the results of the following regression:

$$aver\_diff_i = a + bT_i + cX_i + e_i$$

where $aver\_diff_i$ is the difference in the student's cumulative grade point average for student i, T_i is a dummy variable that takes the value of one if student i is assigned to a joint-liability scheme and zero otherwise[5] and the term X_i is a matrix of controls: year dummy, course dummy, region dummy and age (in months).

Results indicate that students in a joint-liability framework experience a higher increase in the cumulative grade point average than students in the control group. Because of randomization, we assume that if any difference exists between the treatment and control groups in the cumulative grade point average at the end of the experiment, it is due to the effects of joint-liability incentives. Figure 2 presents the differences-in-differences framework and what would have happened to individuals in the treatment group had they not received the intervention. It shows that they should have had a 7.4 cumulative grade point average. However, students in the treatment group averaged 7.8. The difference between the counterfactual and the actual outcome is 0.39 and significant at the one percent level (column 1 of table 6).

[5] Students treated with joint-liability are thus the treatment group of the experiment of the previous year (joint-liability incentives with 8 take-home tests) and the treatment and control groups in the high frequency experiment (joint-liability incentives with 16 or 8 take-home tests).

When we control for macroeconomic course dummy, year dummy, region dummy and age (in months), we find that the effect is still present though slightly different[6]. Finally, controlling for gender, the effect remains essentially the same and is significant at the ten percent level.

We also explore if men and women react differently to the joint liability incentives. We are interested in evaluating the effect of being a female student in a joint-liability framework on the difference in the cumulative grade point average. Therefore, we estimate the following equation and present the results in Table 7:

$$aver_{diff} = a + bT_i + cGender_i + dGender_i * T_i + eX_i + e_i$$

where $Gender_i$ is a dummy variable that takes the value of one when student i is a female student and zero otherwise, and $Gender_i * T_i$ is an interaction term that captures the effect of a joint-liability framework when the student is female. In other words, the effect that a joint-liability framework may have on the difference in the cumulative grade point average is not only limited to b, but also depends on the values of d and $Gender_i$. The term X_i is a matrix of controls: year dummy, course dummy, region dummy and age (in months).

In Table 7, we observe that a joint-liability framework directly increases the difference in the cumulative grade point average by 0.54 for both female and male students. This effect is significant at the one percent level. However, there is a negative effect of the joint-liability framework that arises from the interaction term with gender. When the student is female, there is a decrease in the outcome by 0.68 and this effect is

[6] The only unbalanced control is the year dummy variable.

significant at the ten percent level. Therefore, we test whether the sum of these coefficients -b (0.54) and d(-0.68), coefficients of the variables T_i and $(Gender_i * T_i)$- is different from zero. Taking both effects together, that is, the direct effect of a joint-liability scheme and its interaction effect with gender, we find that a joint-liability framework keeps the academic performance of female students unchanged. However, the performance of male students is increased significantly by 0.54.

Taking these results into account, we find that a joint-liability framework has a positive effect in the difference experienced in the cumulative grade point average in the case of male students. However, it is neutral for female students as their cumulative grade point average remains constant.

Discussion

Although we could not find significant differences between the treatment and control groups in the frequencies experiment, we find that the signs are as expected. This is in line with our hypothesis that high frequency take-home tests improve academic performance through the development of study habits.

In order to contribute to future research, we should point out some lessons learned in order to avoid a faulty design. The lack of significance could be due to the small sample size, which was aggravated by the attrition experienced in the intervention group. Hence, in a future intervention we would increase the sample size. Power limitations prevent us from doing any further analyses of mechanisms or heterogeneous treatment effects and, as a result, many interesting questions regarding who benefits from a higher frequency in incentives or from group versus individual incentives are left unanswered.

Another reason for the lack of significance may be that, although students in the low frequency group had to hand in their tasks every two weeks, they could get a head start by studying with those in the treatment group which would lead to no differences arising (positive spillover effect). In a future intervention, we should design different take-home tests for treatment and control groups in order to avoid spillover effects resulting from treated and control group students studying together.

In the future, it could also be useful to design the high frequency treatment with more variability in order to avoid the possible critique that, in the case that a positive effect of 16 take-home tests against 8 take-home tests is found, it is impossible to disentangle if the crucial issue is the increase in the frequency (regardless of the quantity) or simply the duplication of the frequency.

Also, we should explore the effects of a longer intervention. Extending the high frequency experiment, during a whole academic year (30 weeks), might help us to assess the amount of time necessary to develop strong study habits.

Apart from the question of whether different frequencies alter the experiment, studying how the class and group size as well as prize size may change the results in the joint-liability scheme remains for future research. In a larger class or in a larger group, the costs of monitoring each other may be too high leading students to prefer to simply lose the prize, regardless of the frequency of their take-home tests.

With reference to the external validity of our experiment, the conclusions are limited to similar students in a similar background, that is, freshman students taking introductory courses at university. However, there are potential applications to certain populations which are aware of the benefits of a certain activity but are not capable of changing their behavior without external support. We provide researchers with reliable evidence to apply in a wide array of issues such as performance pay for teachers, lesser use of energy among home-owners, and incentives to employees in a firm. In a future intervention, we plan to include in the experiment students that are more advanced in their undergraduate degree.

In this paper we analyze the impact of high and low frequency take-home tests in a joint-liability framework on six academic outcomes of undergraduate students using a randomized field experiment. We find that there are no significant differences between the treatment and control groups in the high frequency experiment, but that the signs are the expected ones in line with our main hypothesis and with previous literature from the health area: exercising more frequently improves academic outcomes and the mechanism behind this finding might be the study habits developed by exercising.

Our second contribution lies on how incentives designed as a joint-liability scheme have a different impact by gender. Male students improve their academic outcome significantly when placed in a joint-liability framework. However, this incentive design does not result in an improvement in academic performance in females. This is consistent with previous results applying a joint-liability scheme in microfinance for women, but it remains a novel result in the area of education.

A randomized controlled trial is an impact evaluation method that relies on straightforward comparisons of outcomes between treatment and control groups to measure the effects of a program. Thus, a randomized design may provide greater confidence to policymakers because of its simplicity and transparency. Nevertheless, many details in the implementation of an RCT may compromise the evaluation design. Consequently, one of the aims of this paper is precisely to contribute to future research on the evaluation of incentives, to provide researchers with evidence to apply in a wide array of issues (performance pay for teachers, incentives on home-owners to lower-energy use, incentives to employees in a firm), to present lessons to avoid a faulty design and to assess heterogeneous effects such as those associated with gender.

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.

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.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, 128, 1-51.

Bertrand, M., and Pan, J. (2013). The trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics,* 5(1), 32-64.

Cabrera, J.M., and Cid, A. (2013). Joint-Liability vs. Individual Incentives in the Classroom. Lessons from a Field Experiment with Undergraduate Students. Working Paper. Retrieved from: http://www2.um.edu.uy/jmcabrera/Research/Jointly_Liability_vs_Individual_Incentives_28012014.pdf

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

Crepón, B., Devoto, F., Duflo, E., and Perienté, W. (2011). Impact of microcredit in rural areas of Morocco: Evidence from a randomized evaluation. J-PAL working paper. Retrieved from http://economics.mit.edu/files/6659.

Duflo, E., Banerjee, A., Glennerster, R., and Kinnan, C. (2013). The Miracle of Microfinance? Evidence from a randomized evaluation. NBER Working Paper No. 18950.

Fryer, R. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *Quarterly Journal of Economics,* 126, 1755–1798.

Grant, D., and Green, W. (2013). Grades as Incentives. *Empirical Economics,* 44:1563–1592.

Rossi, M. (2014). The impact of individual wealth on posterior political power. *Journal of Economic Behavior & Organization*, 106, 469-480.

Ruigrok, A. N. V., Salimi-Khorshidib, G., Lai, M-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., and Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and Behavioral Reviews*, 39, 34-50.

Table 1 - Definition of baseline characteristics and outcome variables

	Description of the variables	Mean	S.D.	Min	Max	Observations
A) Baseline Characteristics						
Age (in months)	Student's age (in months)	240.743	24.369	218.893	320.712	48
Work	Dummy variable (1= Student works, 0= Student does not work)	0.104	0.309	0	1	48
Volunteering	Dummy variable (1= Volunteer at social activities, 0= otherwise)	0.188	0.394	0	1	48
High school 1	Dummy variable (1= Student attended High School 1, 0= Student did not attend High School 1)	0.167	0.377	0	1	48
High School 2	Dummy variable (1= Student attended High School 2, 0= Student did not attend High School 2)	0.146	0.357	0	1	48
Interior	Dummy variable (1= Student is from the Interior of the country, 0 =Student is from the capital)	0.333	0.476	0	1	48
Hours of sport per week	Hours spent doing sports per week	5.360	3.945	0	15	48
Satisfaction with classmates	Student's satisfaction with classmates. Scale: 1-very unsatisfied, 5-very satisfied.	4.146	0.899	1	5	48
Average grade	Total average grade accumulated in the student's career. (Min=0, Max=12)	7.556	2.103	0	11.2	48
Female	Dummy variable (1=Female, 0=Male)	0.417	0.498	0	1	48
Bachelor in economics	Dummy variable (1= Student is studying for a bachelor in economics, 0=Student is studying for a bachelor in management or accountancy)	0.583	0.498	0	1	48
Travel time to university (minutes)	Minutes spent travelling to university	24.313	18.506	10	120	48
Course	Dummy variable (1= course in Macroeconomics, 0= course in Descriptive Economics)	0.500	0.505	0	1	48
Study in group (in % of the time)	Percentage of time that students study in group	0.335	0.199	0.020	0.82	48
Friends (%)	Percentage of friends in the course	0.125	0.126	0	0.455	48
Still unknown (%)	Percentage of students that are unknown	0.605	0.291	0	1	48

Educational Aspirations	Scale: 1- Bachelor unfinished, 2- Hold a Bachelor's degree, 3-Hold two bachelor's degrees, 4-Hold a master's degree, 5- Hold a Ph.D. degree	3.604	1.106	2	5	48

B) Outcomes at Follow-up						
Grade in midterm exam (standardized)	Standardized grades in midterm exams. (Scale in midterm exams: Min=0, Max=12).	0.000	0.988	-1.876	1.436	42
Average grade of take home-tests (standardized)	Standardized grade of take home-take tests (Scale in take-home tests: Min=o, Max=12).	-0.000	0.989	-2.661	1.529	46
Grade in final exam (standardized)	Standardized grade in final exam (Scale in final exams: Min=0, Max=12).	0.000	0.987	-1.825	2.004	40
Total average grade accumulated in the student's career	Total average grade accumulated in the student's career after the intervention. (Min=0, Max=12)	7.623	1.954	0	10.5	48
Average grade in homework & midterm exams in other simultaneous courses	Average grade in homework & midterm exams in simultaneous courses (not the intervention ones). Min=0, Max=12.	7.615	1.559	4	11.25	47
Average grade in other simultaneous final exams	Average grade in simultaneous final exams (not the intervention courses). Min=0, Max=12.	7.995	1.477	5.25	10.75	47

Table 2 - Pre-treatment characteristics by treatment assignment

.

	Treatment	Control	Difference	Standard error	P-value	Observations
Age (in months)	236.214	245.273	9.059	6.984	0.201	48
Work	0.083	0.125	0.042	0.090	0.645	48
Volunteer	0.167	0.208	0.042	0.115	0.719	48
High school 1	0.125	0.208	0.083	0.109	0.449	48
High School 2	0.167	0.125	-0.042	0.104	0.690	48
Interior	0.375	0.292	-0.083	0.138	0.550	48
Hours of sport per week	5.221	5.500	0.279	1.150	0.809	48
Satisfaction with classmates	4.208	4.083	-0.125	0.262	0.635	48
Average grade	7.879	7.233	-0.646	0.606	0.292	48
Female	0.458	0.375	-0.083	0.145	0.568	48
Bachelor in economics	0.625	0.542	-0.083	0.145	0.568	48
Travel time to university (minutes)	25.291	23.333	-1.958	5.392	0.718	48
Group (1= Macroeconomics, 0 = Descriptive Economics)	0.500	0.500	0.000	0.147	1.000	48
Study in group (in % of the time)	0.337	0.333	-0.004	0.058	0.945	48
Friends (%)	0.142	0.108	-0.035	0.036	0.343	48
Still unknown (%)	0.584	0.625	0.041	0.085	0.627	48
Educational Aspirations	3.708	3.500	-0.208	0.321	0.520	48

Table 3 - The effect of high frequency tasks on academic achievement by outcome

	Treatment	Control	Diff	Standard error	p value	Observations
Effects on educational outcomes						
Grade in midterm exam (standardized)	.2195189	-.2414707	-.4609896	.3002362	.1325528	42
Grade in final exam (standardized)	.0968753	-.1184031	-.2152784	.3158964	.4996976	40
Average grade of take home-tests (standardized)	.0258865	-.0284756	-.0543621	.3088376	.861165	42
Spillover effects						
Average grade in homework & midterm exams in other simultaneous courses	7.958	7.4368	-.5212	.4940708	.2977971	42
Average grade in other simultaneous final exams	8.231591	7.96275	-.2688409	.462439	.5642635	42
Total average grade accumulated in the student's career	8.109091	7.74	-.3690909	.5066478	.4705537	42

Table 4 - Pre-treatment balance by joint-liability treatment

	Joint-Liability	Control group	Difference	Standard Error	p-Value	#Obs.
Age (in months)	238.727	235.150	3.577	-4.478	0.427	87
Female	0.348	0.182	0.166	0.131	0.209	87
Work	0.131	0.182	-0.051	0.110	0.642	87
Volunteering	0.261	0.182	0.079	0.125	0.530	87
Travel time to university (minutes)	27.609	25.455	2.154	-3.959	0.588	87
High School 1	0.174	0.091	0.083	0.102)	0.417	87
High School 2	0.261	0.318	-0.057	0.137	0.678	87
Interior	0.261	0.227	0.034	0.130	0.797	87
Hours of sport per week	3.892	4.955	-1.063	0.948	0.265	87
Satisfaction with classmates	4.130	4.318	-0.188	0.228	0.413	87
Average Grade	7.891	7.627	0.264	0.442	0.552	87
Bachelor in Economics	0.522	0.455	0.067	0.151	0.658	87
Macro Course	0.521	0.545	-0.024	0.151	0.876	87
Study group (in % time)	0.293	0.374	-0.081	0.074	0.274	87
Friends (%)	0.138	0.185	-0.047	0.031	0.127	87
Still unknown (%)	0.547	0.496	0.051	0.071	0.472	87
Educational aspirations	3.826	3.773	0.053	0.269	0.843	87

Table 5 – Baseline Characteristics by gender

	Female	Male	Difference	Standard Error	p-Value	Observations
Age (in months)	235.205	237.501	2.296	3.676	0.534	87.000
Work	0.094	0.109	0.015	0.068	0.823	87.000
Volunteering	0.250	0.164	-0.086	0.089	0.333	87.000
High School 1	0.063	0.218	0.156	0.081	0.058	87.000
High School 2	0.188	0.255	0.067	0.094	0.479	87.000
Interior	0.226	0.255	0.029	0.098	0.769	86.000
Hours of sport per week	3.353	5.809	2.456	0.782	0.002	87.000
Satisfaction with classmates	4.313	4.145	-0.167	0.177	0.347	87.000
Average Grade	8.416	7.476	-0.939	0.351	0.009	87.000
Bachelor in Economics	0.719	0.455	-0.264	0.108	0.017	87.000
Travel time to university (minutes)	27.188	24.400	-2.788	3.716	0.455	87.000
Course	0.313	0.618	0.306	0.107	0.006	87.000
Study in group (in % of the time)	0.345	0.336	-0.009	0.051	0.855	87.000
Friends (%)	0.086	0.185	0.099	0.024	0.000	87.000
Still unknown (%)	0.659	0.478	-0.181	0.056	0.002	87.000
Educational aspirations	3.781	3.636	-0.145	0.226	0.524	87.000
Year	0.625	0.400	-0.225	0.110	0.043	87.000

Table 6 – Effects of Joint-Liability incentives on the difference in student's grade average

	Difference in total average grade accumulated in the student's career		
	(1)	(2)	(3)
Joint Liability Incentives	0.387***	0.401*	0.395*
	[0.231]	[0.230]	[0.233]
Female			0.038
			[0.120]
Year (1=High frequency experiment, 0= Joint-liability vs. individual incentives experiment)	0.356***	0.344***	0.341***
	[0.107]	[0.104]	[0.105]
Group (1= Macroeconomics, 0 = Descriptive Economics)		0.175	0.185
		[0.147]	[0.153]
Age (in months)		-0.001	-0.001
		[0.003]	[0.003]
Region (Interior=1, Montevideo=0)		-0.165	-0.162
		[0.229]	[0.229]
Constant	-0.695***	-0.440	-0.472
	[0.216]	[0.789]	[0.768]
Observations	87	86	86
R-squared	0.197	0.220	0.221

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 7 – Effects of Joint-Liability incentives on the difference in student's grade average considering interaction terms

	Difference in total average grade accumulated in the student's career	
	(1)	(2)
Joint Liability Incentives	0.531**	0.537**
	[0.265]	[0.261]
Female	0.575	0.606*
	[0.394]	[0.338]
Female*Joint Liability Incentives	-0.689*	-0.683*
	[0.409]	[0.361]
Year (1=High frequency experiment, 0= Joint-liability vs. individual incentives experiment)	0.371***	0.355***
	[0.107]	[0.103]
Group (1= Macroeconomics, 0 = Descriptive Economics)		0.195
		[0.148]
Age (in months)		-0.001
		[0.003]
Region (Interior=1, Montevideo=0)		-0.126
		[0.221]
Constant	-0.800***	-0.615
	[0.251]	[0.795]
Observations	87	86
R-squared	0.228	0.251

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Figure 1 - Timeline of the Program and Data Collection

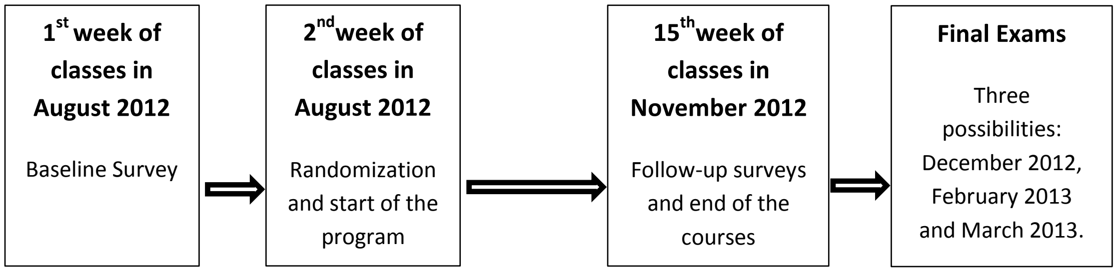
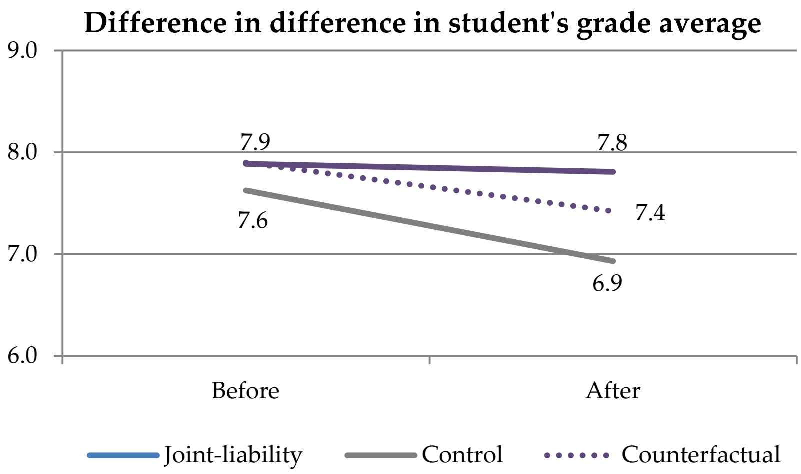


Figure 2


Difference in difference in student's grade average
9.0
8.0
7.0
6.0
7.9
7.8
7.6
7.4
6.9
Before
After
Joint-liability
Control
Counterfactual