Munich Personal RePEc Archive

Gender and high frequency vs low frequency tasks in a context of joint-liability incentives

Cid, Alejandro and Bernatzky, Marianne

Universidad de Montevideo

16 June 2014

Online at https://mpra.ub.uni-muenchen.de/59960/
MPRA Paper No. 59960, posted 21 Feb 2018 14:49 UTC

Gender & High Frequency vs. Low Frequency tasks in a context of Joint-Liability Incentives.

Marianne Bernatzky

José María Cabrera

Alejandro Cid

Universidad de Montevideo *

June 16th, 2014

Abstract

We study the impact of high and low frequency incentives in a joint-liability framework on six academic outcomes of undergraduate students using a randomized field experiment. As recently documented in health literature, incentives to exercise are effective in developing healthy habits. Therefore, we design groups of three students and provide a premium to the homework's grade if all the members of the group (three) meet some requirements. We investigate how the frequency of these take home tests affect the students study habits and thus, the academic outcomes. We find that there are no differences in the student's educational outcomes between the high and low frequency groups.

We also explore if male and female students respond differently to a joint-liability incentives scheme. We find that this treatment improves the accumulated grade average of male students, but not for females. This finding is in line with previous research on joint-liability and gender behavior, but now we present it in a novel context.

Finally, the paper outlines the main evaluation challenges associated with a field experiment in the classroom and provide some lessons to improve evaluation designs and to foster future randomized controlled trials in this area.

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

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

I. Introduction

Recent evidence from the 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 determined number of times during 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 opposed 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 they firstly 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 doing exercise and therefore participants of the experiment engage in more physical activity in the future.

The same relevant question of whether incentives are effective in developing good habits arises in the education field. When we focus on previous literature in 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). Perhaps 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.

However, there is recent evidence that supports joint-liability schemes as a way to provide incentives to students. As the teacher wants to exert effort from his students, he designs a contract in which students 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 fade out 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 if 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 designed groups of three students and provided a premium to the take-home tests' grade if all the members of the group met some requirements. To test how the frequency of these take-home tests could affect the academic performance of the students, two modalities 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 taking into account the findings in the fitness intervention of Charness and Gneezy (2009), we expect that students with high frequency take-home tests will improve their academic performance in comparison to those with low frequency take-home tests, as a result of being more frequently exposed to exercises. Not only we analyze the students' performance in the intervention courses, but also we study the spillover effects 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 (that was exacerbated by the attrition experienced in the evaluation), 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 student's exchange of take-home solutions. These are lessons to improve future field experiments in a similar context.

In the second part of the paper, we explore the impact of a joint-liability framework on the gender gap in the accumulated grade average in the student's career. For this section we pool data from two experiments to increase power. Our results show that the incentives designed as a joint-liability scheme improves students performance. But this overall positive effect masks gender disparities, since it is concentrated in male students who improve their academic outcomes significantly. However, 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) that 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, and section IV the discussion and V conclusion.

II. 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 has 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. We have a sample of 48 students over 18 years old with a mean average grade of 7.5; two thirds are from the interior of the country, nearly 42% are female students 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% are social volunteers 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 behaviors, students of the sample devoted 33% of their time of study to do so in groups. When students were asked about the share of classmates that were friends, we find that, on average, 13% of the classmates are friends. Analogous, the average percentage of unknown students is 60%.

[Insert Table 1]

The aim of the intervention is to test if high frequency take-home tests improve or not academic performance under a joint-liability scheme. We designed a randomized trial in order to evaluate the intervention, with the approval of the ethical review board of the university.

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. They were assigned 16 take-home tests during 15 weeks of classes.

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

For both treatment and control group, take-home tests did not require team work. Each student was required to hand in the solutions in a personal sheet at the beginning of the class (they were allowed to prepare the take-homes with another classmate). It is important to notice that the content of the 16 take-home tests is exactly the same to the content of the 8 take-home tests; the difference between groups is in the distribution of tasks in the 15 weeks of classes. So, the treatment is a variation only in the *frequency* of the exercise, not in the *total amount* of exercise performed in the course.

There were 48 students in this field experiment: 24 in Macroeconomics and 24 in Descriptive Economics. 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, travel time to university, academic expectations and number of friends in the course. 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.

Timeline of the Program and Data Collection

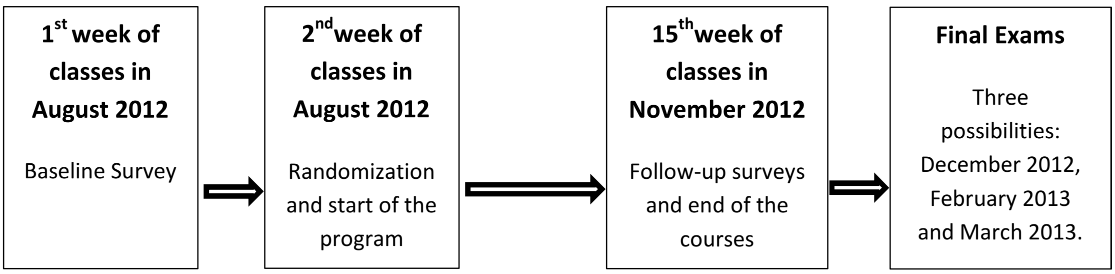

Table 2 shows that the three groups had similar characteristics. They were balanced in eighteen observable variables.

[Insert Table 2]

It is necessary to mention that some observations suffered attrition. 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 in Treatment group and 4 students in 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 into account this information, we compared pre-treatment characteristics of the individuals that suffered attrition and the students that remained in the treatment/control groups: all the variables remained balanced[1].

III. Econometric Model and Results

High Frequency vs. Low Frequency Take- Home Tests

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[2], 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)[3], $T_i$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

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

[2] 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.

[3]In Panel B of Table 1 we present a description of the outcome variables at the follow-up.

Macroeconomics course and zero otherwise, X_i is a matrix of student characteristics, and e_i is the error term. Given the random assignment to the treatment and 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 raise 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 treatment on *midterm exams* only if the differences are taken at the fifteen percent level.

[Insert Table 3]

In addition to this, it could be argue d that the high frequency of the take-home tests on the experimental courses may have worsen 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, total average grade in the student's career*. 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 the 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 the attrition in the intervention: six students in our sample is a relevant share. Another reason is that, although students in the low frequency group had to hand in their tasks every two weeks, they could advance their tasks studying with those of the treatment group (positive spillover effects) and therefore no differences in habit formation arise. In this sense, the spillover effects are present. Also, it could happen that the difference in the frequency of take-home tests is too small to see differences in effort among students; or that the duration of the intervention (15 weeks) is too short to see changes in habit formation. Finally, we do not rule out the possibility that what is only working is the joint-liability incentive and the frequency is not a relevant factor. We properly discuss each of these arguments in the last section of this paper.

<u>Gender & Academic Outcomes</u>

Gender differences have been widely documented. Boys have more attention and behavioral difficulties, lower levels of inhibitory control and perceptual sensitivity and are more likely to be diagnosed attention deficit hyperactivity disorder (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 structure in areas related to mood, emotions and emotion regulation. Social influences may be related to home and school environment (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 accumulated grade 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 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 accumulated grade average on female students.

For this section we make use of an experiment conducted the year before the current experiment. We will take advantage that we designed the two experiments with the same joint-liability framework, they were implemented in the same courses and with similar populations, they share the same baseline survey, and we have a homogeneous outcome in 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 new experiment presented in the first part of the paper the interest was in 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. In table 4 we present the balance in 15 pre-treatment variables for this experiment (that pools data from two years). Students treated with joint-liability incentives are in fact identical to students from the control group. There is only one variable with a slightly significant difference (p-val 0.095), but 1 variable in 15 that is significant at the 10% level is less than what one would expect by random chance. The only difference between the pool of subjects 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 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 the country, they have similar educational aspirations, devote a similar share of their time to study in group, to travel to university and to volunteer in social activities and the percentage of them that have a job is alike. Also, the satisfaction with classmates is akin. However, male students tend to spend more time practicing sports, have more friends in the classroom and less classmates considered unknown. Female students have a higher average grade – almost one point of difference: 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. In the second year of the experiment there were more women participating.

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

Taking this into account, we first report the effect of being in a joint-liability framework on the difference experienced in the accumulated grade 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 first regression, expressed in the following equation:

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

where $aver\_diff_i$ is the difference in the student's accumulated grade 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[4].

[Insert Table 6]

Results indicate that students in a joint-liability framework experience a higher increase in the accumulated grade 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 accumulated grade average at the end of the experiment, it is due to the effects of joint-liability incentives. Figure 1 shows the differences-in-differences framework and what would have happened to individuals in the treatment group if they had not received the intervention. They should have 7.2 as their accumulated grade average. However, students in the treatment group averaged 7.8. The difference

[4] 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).

between the counterfactual and the actual outcome is 0.62 and significant at the one percent level (column 1 of table 6).

When we control for macroeconomic course dummy, year dummy, region dummy and age (in months), we find that the effect is still present and was slightly affected[5]. Finally, controlling for gender, the effect remains essentially the same and is significant at the ten percent level.

We also estimate if men and women react in different ways to the joint liability incentives. We are interested in considering the effect of being a female student in a joint-liability framework on the difference in the accumulated grade 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 accumulated grade 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).

[Insert Table 7]

We observe that a joint-liability framework directly increases the difference in the accumulated grade's 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-

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

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 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. Taken together both effects, 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 accumulated grade average in the case of male students. However, this framework prevents female students from performing better and therefore their accumulated grade average remains constant.

IV. Discussion

Although we could not find significant differences between the treatment and control groups in the frequencies experiment, we find that the signs are the expected ones. 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 lessons in order to avoid a broken design. The lack of significance could be due to the small sample size, which was aggravated with the attrition of the intervention. In a future intervention we should 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 about 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 advance their tasks studying with those of the treatment group and therefore no differences arise (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 arising from treated and control students studying together.

In next interventions it could be useful to design the high frequency treatment with more variability in order to avoid the possible critic that in case of finding positive effects of 16 take-home tests against 8 take-home tests, it is impossible to disentangle if the crucial issue is the increase in the frequency in any amount 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 necessity of time to develop strong study habits.

Apart from questioning whether different frequencies may 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 that students simply prefer to lose the prize, no matter 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 as long as a certain population is aware of the benefits of a determined activity but is not capable of changing its behavior without external support. We provide researchers with reliable evidence to apply in a wide array of issues such as performance pay for teachers, home-owners to use less-energy, incentives to employees in a firm. In a future intervention, we plan to include in the experiment students that are advanced in their undergraduate courses.

V. Conclusions

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. There are no significant differences between the treatment and control groups in the high frequency experiment, but we find that the signs are the expected ones in line with our main hypothesis and with previous literature in the area of health: exercising more frequently improves academic outcomes and the mechanism behind this finding might be the study habits developed by exercising.

Our second contribution is to show that incentives designed as a joint-liability scheme have an different impact by gender. Male students improve their academic outcome significantly when placed in a joint-liability framework. But this incentives design does not help female students to improve their academic performance. This is consistent with previous results applying a joint-liability scheme in microfinance for women, but this is 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, home-owners to use less-energy, incentives to employees in a firm), to present lessons to avoid a broken design and to assess heterogeneous effects such as gender ones.

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.

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.

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)	237.166	235.150	-2.017	4.082	0.623	87
Work	0.077	0.182	0.105	0.075	0.166	87
Volunteering	0.200	0.182	-0.018	0.099	0.855	87
High School 1	0.185	0.091	-0.094	0.091	0.307	87
High School 2	0.200	0.318	0.118	0.104	0.260	87
Interior	0.262	0.227	-0.034	0.108	0.753	87
Hours of sport per week	4.889	4.955	0.065	0.917	0.943	87
Satisfaction with classmates	4.169	4.318	0.149	0.196	0.450	87
Average Grade	7.888	7.627	-0.260	0.405	0.522	87
Bachelor in Economics	0.585	0.455	-0.130	0.123	0.294	87
Macro Course	0.492	0.545	0.053	0.125	0.671	87
Study group (in % time)	0.327	0.374	0.047	0.056	0.408	87
Friends (%)	0.136	0.185	0.048	0.029	0.095	87
Still unknown (%)	0.561	0.496	-0.065	0.065	0.318	87
Educational aspirations	3.662	3.773	0.111	0.252	0.660	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	0.617***	0.401*	0.395*
	[0.222]	[0.230]	[0.233]
Female			0.038
			[0.120]
Year (1=High frequency experiment, 0= Joint-liability vs. individual incentives experiment)		0.344***	0.341***
		[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
	[0.214]	[0.789]	[0.768]
Observations	87	86	86
R-squared	0.152	0.220	0.221

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

Figure 1

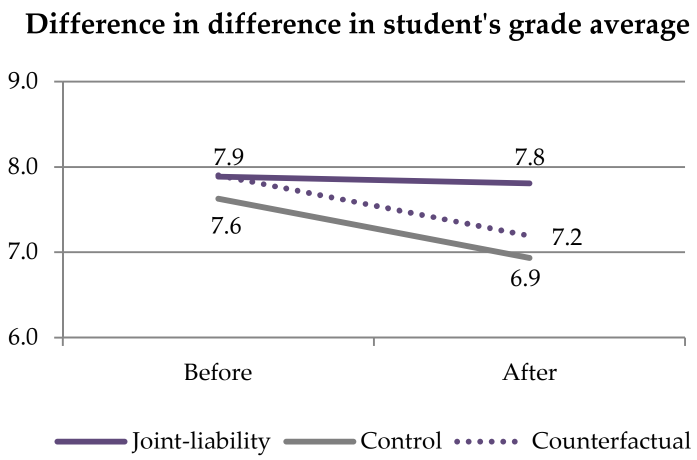

Table 7 – Effects of Joint-Liability incentives on the difference in student's grade average considering interaction terms

	Difference in total average grade accumulated in the student's career	
	(1)	(2)
Joint Liability Incentives	0.751***	0.537**
	[0.259]	[0.261]
Female	0.575	0.606*
	[0.391]	[0.338]
Female*Joint Liability Incentives	-0.644	-0.683*
	[0.409]	[0.361]
Year (1=High frequency experiment, 0= Joint-liability vs. individual incentives experiment)		0.355***
		[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
	[0.250]	[0.795]
Observations	87	86
R-squared	0.180	0.251

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