Journal of Behavioral Education
https://doi.org/10.1007/s10864-023-09524-6

ORIGINAL PAPER

The Use of Innovative Incentives in the Classroom to Explore the Impact of Peer Monitoring on Academic Achievements

Alejandro Cid[1] · **José María Cabrera**[1,2]

Accepted: 23 July 2023

Abstract

In a population of undergraduate students, we examined the impact of reciprocal peer monitoring of educational behaviors on academic performance. Reciprocal peer monitoring is a novel design of incentives that promotes peer-observing and checking the behavior of others. To distinguish the pure effect of peer monitoring from self-motivation, we also examined the effects of individual incentives on academic performance. Using a randomized controlled trial, this study showed that a joint-liability incentives arrangement was more effective than the individual incentives approach to increase students' academic performance. The results also showed that participants reported negative views of aspects of the joint-liability incentives intervention. The current procedures entailed a novel system of incentives for students that does not require tangible reinforcers and requires them to exert more effort to succeed in a course. These procedures may be characterized as an innovative insight for the design of grading policies in the classroom and other social settings.

Keywords Field experiment · Randomization · Education · Joint liability · Student incentives · Reinforcers · Behavioral economics

Introduction

Incentives for teachers for improving their performance (develop new skills, increase long-run effort, etc.) have received considerable attention in previous literature (Duflo et al., 2011; Fryer, 2013; Fryer et al., 2012; Speroni et al., 2020) but less attention has been paid to encouraging students; and the literature is not

✉ Alejandro Cid
acid@um.edu.uy

[1] Universidad de Montevideo, 2544 Prudencio de Pena St., 11600 Montevideo, Uruguay

[2] Universidad de Montevideo, Montevideo, Uruguay

Published online: 09 August 2023

conclusive (Angrist et al., 2009, 2014; Fryer, 2011; Grant & Green, 2013; Rassuli, 2012). Bishop, (2006) observed that student effort and engagement vary substantially within and across schools, and teachers need to encourage student cooperation if educational goals are to be achieved; classroom goals are often negotiated between teacher and students, and the behavior of these system's agents depend on the incentives that exist within that context.

Behavior change occurs when certain consequences are contingent upon performance. "A consequence is contingent when it is delivered only after the target behavior has been performed and is otherwise not available" (Kazdin, 2012, p. 61). Teachers may design positive reinforcers to foster some responses. Positive reinforcement refers to the increase in the likelihood of a response (under similar circumstances) when followed by a favorable consequence.

Grades designed as individual incentives, or even tangible reinforcers (e.g., tokens, paper coupons, edibles), 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 & Green, 2013). With regard to homework, for example, time devoted to the activity varies a considerably (Bishop, 2006). Thus, teachers often need to incentivize student effort. It is a common practice to provide students with feedback along with marks for assessed work, but there is some debate over whether students make effective use of it (Holme and Forshaw, 2009; Price et al., 2010). For instance, students may ignore feedback due to a lack of time to reflect, or because they are extrinsically motivated and only concerned with formal achievements (Holme & Forshaw, 2009).

Peer monitoring by classmates may increase the likelihood that target students will consider teachers' feedback. Brown et al., (1999) defined peer monitoring as the observation and checking of others' behavior within a group with a focus on appropriateness (i.e., meet their commitments in the agreed way) and effectiveness (i.e., complete, in time, what they set out to do). Recent research has explored the effects of peer monitoring by classmates on educational achievement. Findings to date are promising (e.g., Larsen et al., 2020) but most interventions have been applied only among students with learning disabilities and emotional-behavioral difficulties (e.g., Davies & Witte, 2000; Morrison et al., 2001; Smith et al., 2015).

This study provides a novel system of incentives for students that does not require tangible reinforcers and requires more effort to succeed in the course. The idea is borrowed from the microfinance literature. In the microfinance framework of microcredit experience, the key to success for the repayment rate is considered to be the joint-liability mechanism. Said another way, 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 make a good peer selection and monitor each other (Banerjee & Duflo, 2010; Becchetti & Pisani, 2010). Thus, in the microfinance literature, peer selection is considered an important part of the explanation of why joint liability may generate positive effects. Peer monitoring is another and the current study contributes to previous literature by isolating it as a mechanism. We designed a joint-liability contract in the classroom that randomized

students into groups and gave students strong incentives to monitor each other. Peer monitoring by the other group-mates exerted pressure on students to comply with the requirements of the course for achieving the grade premium. Eventually, due to peer monitoring, the students devoted more effort to study resulting in better grades and academic achievements. Precisely, the present study aimed to identify the impact of providing joint-liability incentives to undergraduate students on academic achievements. An additional research objective was to explore the channels that may be operating between this novel structure of incentives and students' academic performance. Hence, we explored the effects of joint-liability incentives on academic achievements guided by the following research questions including whether or not (a) joint-liability incentives impact students´ grades at the course's assignments; (b) innovative incentives have spillover effects on the other simultaneous courses attended by the experimental group; and (c) the joint-liability incentives have effects on the academic performance of students?

Method

Field Experiment

In this randomized controlled trial, the individuals who satisfy the eligibility requirements for the new program of incentives (i.e., undergraduate students of two courses—Macroeconomics I and Descriptive Economics-that accept to participate) were randomly divided into the Treatment 1 Group (joint-liability incentives), Treatment *2* Group (individual incentives), and Control Group (no incentives). Both the instructor of Macroeconomics I and the instructor of Descriptive Economics divided approximately 40% of their students into subgroups of three students and offered a premium to their homework grade that required that all three members of the subgroup achieved some goals (the Treatment 1 Group).

To distinguish the pure effect of peer monitoring from self-motivation, another 30% of the students of the two courses were randomly assigned to receive individual incentives (the Treatment 2 Group). Finally, the rest of the students of both courses were assigned to the Control Group (they would not receive any incentives).

Hence, applying randomization, students were assigned to the joint-liability treatment, or to the individual incentives treatment, or to the control group. Because assignment is random, the treatment 1, treatment 2, and control groups would be expected to have similar outcomes in the post-program period in the absence of the program intervention. Randomization, therefore, provides a simple method for constructing a counterfactual for the treatments group. The estimate of the program effect from a randomized evaluation is simply the difference in post-program outcomes between treatment 1 group, treatment 2 group, and the control group.

Additionally, considering that the designed incentives might introduce socioemotional challenges (i.e., some students may feel uneasy due to peer monitoring or due to some kind of hidden competence that could arise within classmates, for

instance), we tested the impact of the intervention on the students' satisfaction with their classmates.

Participants

There were 51 different students in this field experiment: 26 in Macroeconomics I and 25 in Descriptive Economics. In the first week of the course, all 51 applicants were asked to take part in a survey. Survey data were collected on a wide array of student characteristics including age, gender, working hours, hours devoted to sports and volunteering, high school of origin, region of the country they come from, the distance between their home in Montevideo and the university, academic expectations, and the number of friends in the course. The university also provided other demographic data such as average grade in previous courses and the number of credits already earned at the university. Students were asked if they wanted to participate in a pilot project where some of them would receive a bonus to their grades, subject to some requirements. To avoid any distress in those that would not receive the bonus in the experiment, the researcher emphasized that no one would be harmed in their grades in any way, that it was a pilot project, and that they would be provided with similar opportunities in future. All of them agreed to participate in the experiment.

Setting

The experimental courses were core classes for undergraduate students in their first year at Universidad de Montevideo, a private university in Uruguay—a developing country in Latin America. The course composition was primarily students majoring in Economics and Business. Undergraduate students at Universidad de Montevideo have to complete some credits (one credit corresponds to ten hours of class) in core courses to obtain their bachelor's degree. 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: a midterm exam (35% of the final grade), eight take-home tests (15%), and a final exam (50%). The frequency of take-home tests was approximately one every two weeks. The final exam covered the same material as the midterms and take-home assignments (that is, the students do not need to study additional material for the final exam). The minimum grade to pass the course was 6 in a scale from 1 to 12. Also, attendance to class was mandatory. Each course had 60 class meetings of 50 min each distributed throughout 15 weeks. Students were allowed up to 15 absences. There was nothing atypical about the characteristics of these courses or the grading system in comparison with other courses offered at Universidad de Montevideo. For instance, in most courses at the university, while it is rare that a student shows more than 15 absences (and thus has to dropout from the course), it is also unusual that students attend all the 60 class meetings. Specifically, the current experimental joint-liability program

consisted of giving incentives for attendance in addition to stimulating take-home tests.

Procedures

We faced two major challenges to determine causal effects. The first one entailed self-virtuous group selection (to minimize the probability of losing the reward, no one wanted to be grouped with classmates that show poor academic performance), which was overcome with the random assignment of participants into groups. A second challenge was, that if faced with an incentive, an individual may exert more effort, whether she/he is in a group or not. If the study only identified the effect of the joint-liability treatment in comparison with the control group, we would not have been able to distinguish between the reinforcing effects of the prize and the effects of the peer monitoring component, because the joint-liability treatment includes both components (prize and peer monitoring). Thus, we constructed two different conditions including (a) individual and joint-liability incentives and (b) control. With the current design, we sought to isolate the effects of the monitoring effect of peers.

As Fig. 1 shows, in the second week of the course, students were randomly distributed into three groups, stratified by course. In all three groups, take-home tests did not require teamwork, including joint-liability group. Students could solve the problem sets alone, with members of their team, or with students from one of the other two group. Each student was required to hand in her personal sheet with solutions at the beginning of the class. Solutions could be identical across students.

Because the number of students in the joint-liability group had to be a multiple of three, 24 students were randomly assigned to the joint-liability group 1, 14 were assigned to the individual incentive group, and the remaining 13 participants were assigned to the control group. Following random allocation, we conducted t test to check the balancing of the groups. If there were significant differences (i.e., 10% or above) in mean pre-treatment characteristics between the control and treatment groups, we repeated the random assignment procedure until sufficient balancing was obtained.

Joint-Liability Condition

In the joint-liability group (Treatment group 1), students were randomly assigned to a group of three and received a 20% increase in the grade of each take-home test if

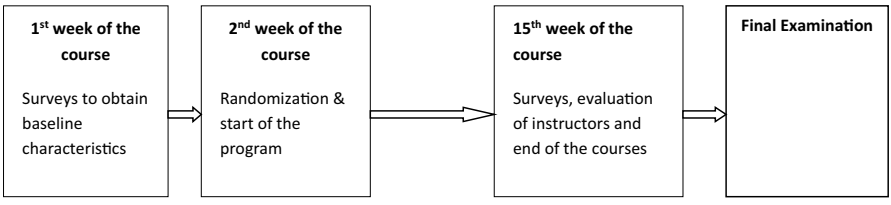

Fig. 1 Timeline of the program and data collection

each student of her group fulfilled two conditions including (a) a grade of at least 6 in the take-home test and (b) no absences during the week in which the take-home test had to be handed in.

Individual Incentive Condition

The procedures for the individual incentive group (Treatment group 2) were similar to the joint-liability group except that grades did not depend on the compliance of others. Specifically, the student received a 20% increase in the grade of each take-home test if she/he (a) obtained a grade of at least 6 in the take-home test and (b) had no absences that week.

Control Condition

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

Data Collection at Follow-up

In the final week of the course (i.e., the 15th week; see Fig. 1), the instructors provided the data on percentage of take-home tests handed in, average grade in take-home tests, and grade in midterm exam. After the examination period, the university provided administrative data pertaining to 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, and total number of credits earned in the semester. Research assistants graded take-home tests and were blinded to the distribution of students among the different treatments.

On the last day of the course, we surveyed both treated and control students about their satisfaction with their classmates—we had the hypothesis that peer monitoring may have distorted the socioemotional environment of the class. In this self-administrated survey, we also included a question about the satisfaction with the neighborhood where the university is placed, in order to use this variable to run a placebo test. Because there is no plausible process by which the program could affect the students' satisfaction with the neighborhood where the university is located, the analysis should have found negligible effects on the outcome.

Data Analysis

Table 1 reports group balancing and shows that the three groups had similar characteristics. They were balanced based on 18 observable variables. By the random allocation design, the probability of receiving treatment was orthogonal to students' characteristics. Therefore, including these characteristics in the regression model,

Table 1 Pre-treatment characteristics by treatment assignment

	Joint liability	Individual incentive	Control	Diff (JL-II)	Diff (Control-JL)	Diff (Control-II)
Age (in months)	238.904	233.757	237.605	−5.147 (5.165)	−1.298 (5.614)	3.848 (4.576)
Male	0.666	0.785	0.846	0.119 (0.155)	0.179 (0.154)	0.060 (0.155)
Average grade	7.970	7.328	7.453	−0.642 (0.546)	−0.516 (0.521)	0.125 (0.528)
Credits earned	53.333	35.642	48.423	−17.690 (15.549)	−4.910 (18.157)	12.780 (15.015)
Bachelor in economics	0.541	0.500	0.538	−0.041 (0.172)	−0.003 (0.176)	0.038 (0.199)
Work	0.166	0.214	0.076	0.047 (0.133)	−0.089 (0.120)	−0.137 (0.139)
Volunteering	0.250	0.214	0.153	−0.035 (0.146)	−0.096 (0.144)	−0.060 (0.155)
Interior*	0.250	0.357	0.307	0.107 (0.155)	0.057 (0.156)	−0.049 (0.188)
High School 1**	0.291	0.285	0.230	−0.005 (0.156)	−0.060 (0.156)	−0.054 (0.175)
High School 2	0.166	0.071	0.076	−0.095 (0.115)	−0.089 (0.120)	0.005 (0.104)
Hours of sports per week	3.812	5.178	4.423	1.366 (1.095)	0.610 (1.051)	−0.755 (1.185)
Satisfaction with classmates***	4.166	4.214	4.307	0.047 (0.272)	0.141 (0.260)	0.093 (0.318)
Travel time to university (in minutes)	27.708	27.142	22.692	−0.565 (4.667)	−5.016 (4.649)	−4.450 (3.786)
Group (1 = Macroeconomics; 2 = Descriptive Economics)	1.500	1.500	1.461	0.000 (0.172)	−0.038 (0.176)	−0.038 (0.199)
Study in group (in % of the time)	0.280	0.350	0.411	0.069 (0.078)	0.131 (0.085)	0.061 (0.094)
Friends in the classroom (%)	0.133	0.184	0.119	0.051 (0.036)	−0.013 (0.036)	−0.064 (0.042)
Still unknown (%)	0.557	0.500	0.588	−0.056 (0.077)	0.030 (0.084)	0.087 (0.095)
Educational aspirations****	3.875	4.000	3.461	0.125 (0.320)	−0.413 (0.318)	−0.538 (0.386)
Observations	24	14	13			

Standard errors in parentheses

JL Joint liability, *II* Individual incentive

*Interior is a dummy variable that takes the value 1 if the student attended a high school that was not located in the capital city of the country

**High School 1 and high school 2 take the value 1 if the student attended the most prestigious high schools in terms of academic achievements

***Satisfaction with classmates is a variable that reports the answer to the question please, on a scale from 1 to 5, indicate how satisfied are you with your classmates, taking into account that 5 indicates the highest satisfaction

****Educational aspirations is a variable that reports—on a scale from 1 (dropout from college) to 5 (Ph.D.)—the maximum educational level that the student aspires to complete

while it might have reduced standard errors, was not necessary for consistency. Our findings do not change if controls are included in the estimates.

As is typical in studies that follow students during the duration of classes, some attrition occurred. In the last week of the course, two participants from joint-liability group, one from individual incentive group, and three from the control group dropped out of the program. We had collected some outcome data from the six participants who had dropped out of the study during the courses and from follow-up administrative data. However, we were not able to collect complete outcome data (i.e., grade in midterm exam, satisfaction with classmates, evaluation of the instructor) from them for several reasons (e.g., most were freshmen and it is typical of freshmen to change to other degrees early in their study; some dropped out of the course before the midterm exam; some refused to evaluate the instructor because the evaluation demanded extra time out of class).

The pre-treatment characteristics of the individuals that dropped out of the study were compared, via a t test on the equality of means, with those students who remained in the treatment and control groups. Because 15 out of the 18 variables remained balanced, the baseline data provided a measure of the similarity between the two groups. Only three variables were not balanced: (a) a dummy variable that indicates that the student came from the interior of the country, (b) the percentage of friends in the classroom at the beginning of the course, and (c) the percentage of classmates that were strangers at the beginning of the course. These variables were included in the regressions as a robustness check, and our findings were not modified by adding these controls to the regressions.

Results

The primary purpose of this study was to determine the potential relative effects of joint-liability incentives for undergraduate students and individual incentives on students' achievements. Formally, the study aimed to estimate the following equation:

$$Y_i = a + bT1_i + cT2_i + d\text{Group}_i + X_i'f + e_i \quad (1)$$

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 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 the inexistence of no-compliers, this equation can be estimated consistently with Ordinary Least Squares (OLS).

Table 2 The effect of incentives on academic achievement

	Dependent variable: index of academic achievement	
	(1)	(2)
Joint liability	0.460** (0.202)	0.437* (0.218)
Individual	0.189 (0.225)	0.165 (0.241)
Controls		
Gender	No	Yes
Age	No	Yes
Observations	43	43

Robust standard errors in parentheses

All models control by the course taken by students (Macroeconomics or Descriptive Economics)

This table considers 43 individuals because, 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)

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

*Significant at the 10% level

**Significant at the 5% level

Impact on an Aggregated Index of Educational Attainment

To draw general conclusions in the context of multiple outcomes, Table 2 shows the findings of a summary index that aggregates information over the eight educational outcomes. The construction of this summary index followed the procedure described by 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 scores. Z scores are calculated by subtracting the control group mean and dividing by the control group standard deviation. 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.

The effect of the joint-liability intervention on the overall index that averages together all eight outcomes was statistically significant with a size of overall effect around 0.45 standard deviations, in comparison with the control group (see Table 2). These results were similar when controlling for the variables that were unbalanced due to attrition (interior as region of origin, number of friends in the class, number of totally unknown people in the class). Given that grades on take-home tests in the

experimental courses may not reflect the real effort deployed by the students (e.g., students may copy the answers due to the pressure exerted by peer monitoring), the researcher also built the index without the variable average grade in take-home tests and the results were similar. This positive average effect of the joint-liability mechanism has also been shown to be present in other research areas such as microfinance (Banerjee & Duflo, 2010; Becchetti & Pisani, 2010).

Alternatively, the individual incentives arrangement had no significant effect on the students' performance in the course (see Table 2). This result is in line with previous findings that suggest that though grades may theoretically be valuable as an ability signal in the job market (Zubrickas, 2013), they may not be effective motivators in classes at universities (Grant & Green, 2013), at least when they are designed as individual incentives.

Impact on Each Educational Outcome

The fact that the joint-liability incentives arrangement increased the index of overall performance may be the result of different patterns of effects over the individual outcomes that were included in the index. Thus, Table 3 reports the effect of the treatments on each of the eight educational outcomes that were linked to the students' academic performance. The results were similar when no controls were included and when the variables that were unbalanced due to attrition were control for.

The first column of Table 3 reports the effects on the percentage of *take-home tests handed in* by the students. The joint-liability incentives intervention appeared to have positively impacted the homework completed by students. Specifically, the percentage of completed take-home tests was 18% which was 30% higher than what was demonstrated by the control group. The individual incentives intervention did not show any significant impact. The second column shows the effect of the respective interventions on the average grade of take-home tests (the 20% prize in this average grade was not included). The results of the average grade in take-home tests were standardized for each of the courses (Macroeconomics and Descriptive Economics). The standardized grades were calculated by subtracting the course mean (Macroeconomics I or Descriptive Economics) and dividing by the resulting number by the course standard deviation.

While the joint-liability incentives intervention increased the standardized average grade in take-home tests by 0.64, the individual incentives intervention seemed to have no effect. The third column shows us the impact of the treatments on midterm examinations. Also, the results of grades in midterm exams for each of the courses were standardized (Macroeconomics and Descriptive Economics). Those who received the joint-liability incentives interventions outperformed the control group by nearly 0.7 in the standardized grades of midterm examinations. The individual incentives intervention did not show a significant impact. Column 4 shows that the estimates did not report a significant impact on the grade in the final examination. At first sight, these findings could suggest that the positive impact of group

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
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)
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

incentives is present only in the short-run (i.e., a higher percentage of take-home tests handed in with higher grades on average and higher grades in midterm exams) and fades in the long run (i.e., there is no improvement in the grade in the final exam among students who receive the treatments). Moreover, it may be stated that extra incentives may distort the amount of time that students devote to different courses during the semester. In other words, extra incentives may divert the students' efforts from other courses, resulting in poorer results in the grades achieved in other courses. The possibility of negative spillover effects is also explored in Table 3.

The 5th column of Table 3 reports the effects of the interventions on the average grade achieved in midterm exams and homework of other simultaneous courses taken by the students in the same semester. The individual incentives intervention did not show a significant impact; however, the joint-liability incentives intervention increased the average grade of midterm exams and homework of simultaneous courses by 1.15 (i.e., an increase of nearly 20% in comparison with the control group).

Though column 6 of Table 3 reports no improvement in the average grade in the final exams of the other simultaneous courses among the students who received the treatments, the 7th and 8th columns show positive spillover effects. The joint-liability incentives intervention increased the accumulated grade average attained by the students in their undergraduate courses by nearly 0.8 (i.e., a 12% increase relative to the control group). The joint-liability incentive intervention also increased the credits earned in the semester by (i.e., an increase of more than 20% relative to the control group). Hence, the joint-liability incentives intervention increased the student's overall academic performance in the semester.

Table 4 The effects of incentives on satisfaction with classmates

	Dependent variable: index of satisfaction with classmates		
	(1)	(2)	(3)
Joint-liability incentives	−0.474 (0.293)	−0.513* (0.258)	−0.502* (0.278)
Individual incentives	−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

Effect on Subjective Well-being

Exploiting the data available from the follow-up survey, we sought to measure whether the joint-liability incentives intervention effected the students' subjective well-being. As Table 4 shows, group incentives negatively impacted students' satisfaction with classmates the results were similar when controlling for the variables that are unbalanced due to attrition. This effect may have been impacted by the fact that the students who received the joint-liability incentives intervention were assigned to groups of three students by randomization. That is, to win the prize of an extra 20%, each one in the group of three needed to fulfill the requirements (i.e., attendance to class and minimum grade on take-home tests). If one of the three classmates of the group did not fulfill the requisites, all of them lost the prize, despite the individual effort made. Many of these students were freshmen from different high schools of origin and they were not necessarily close friends; however, they were required to interact within a group. At times they may develop some reproaches toward the other members of the group. For instance, every time one of them did not hand in the homework, she made the other members of the group lose their prize. But on 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. Evidence of negative implications of peer pressure is present in previous literature. Montgomery, Bhattacharya, and Hulme (1996) noted examples of Bangladesh Rural Advancement Committee (BRAC) group members taking aggressive action against defaulters (seizing the individual's household goods or tearing down a woman's house because she had not complied with the groups' goals). This drawback should be taken into account also in the design of joint-liabilities schemes in other contexts (teachers, workers in a company, etc.).

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

	Joint liability	Individual Incentive	Control	Diff (JI-II)	Diff (control-JL)	Diff (control-II)
Mean	0.239	−0.534	0.139	−0.773** (0.347)	−0.100 (0.339)	0.673 (0.439)
Observations	21	12	10			

Standard errors in parentheses

Given that the evaluation of professors is confidential information, it was not possible to use individual-level data; instead, aggregated data from the evaluation of professor—by the joint-liability intervention, the individual incentive intervention, and control group—were provided by the IT department of the university

JL Joint liability, *II* Individual incentive

** Significant at the 5% level

Effect on How Students Evaluate the Instructor

As Table 5 reports, students who received the joint-liability incentive intervention did not seem to impact evaluation of the instructor. However, the individual incentives intervention did appear to negatively the evaluation of the instructor of the course. Gneezy et al. (2011) provided a possible explanation by suggesting 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 arrangement makes it clear to the rest of the classmates whether the student has met the requirement or not, because both the assignments to treatments and control groups and the prizes awarded were available to all the students of the course. Being in the individual incentives group and obtaining the prize could be seen as an unattractive signal within classmates, concluding in a lower personal image (i.e., an individualistic person). Thus, the student might have ended up unhappy with the instructor for being assigned to the individual incentives treatment.

A common concern in the evaluation of programs by randomization is that results from the control group may be negatively affected in terms of motivation suppression (i.e., because extra incentives are not available as they are for the incentive groups). However, Table 5 reports 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.

Table 6 False experiment—satisfaction with the university's neighborhood

	(1)	(2)	(3)
Joint liability	−0.367 (0.366)	−0.317 (0.366)	−0.408 (0.375)
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)

Robustness Check about the Impact on Satisfaction

The presence of some unbalanced attrition in the follow-up survey might have undermined the findings on satisfaction. To assess the possible effects of unbalanced attrition, we estimated the average causal effect running a placebo test. We hypothesized that the program would not affect the students' satisfaction with the neighborhood where the university was located. As expected, Table 6 shows that the evaluation found no significant impact of the joint-liability incentive intervention on the students' satisfaction with the university's neighborhood (i.e., the results were similar when controlling for the variables that were unbalanced due to attrition). Thus, it may be inferred that the previous finding about satisfaction with classmates was operating through the joint-liability mechanism and was not a spurious correlation.

Discussion

Prior research has suggested that graded homework causes students to devote more effort than when they are assigned non-graded homework (Pozo & Stull, 2006). The current study sought to evaluate whether providing joint-liability and individual extra incentives for take-home tests raise the student's overall academic performance; and what are the relative effects of the two incentive arrangements.

Previous literature has stressed the informational and enforcement advantages of joint-liability contracts. Joint liability encourages peer monitoring to reduce moral hazard (i.e., the reduction of the probability that someone enters into a contract in the absence of good faith) and provides groups' members with incentives to enforce the commitment to the rules (Ghatak & Guinnane, 1999). However, a major obstacle to positive effects of joint liability may arise when social ties among members of the group are too weak to support feelings of group solidarity. Joint liability also may not be effective if the individuals involved are unwilling, for whatever reason, to put pressure and sanction those who default.

Our study shows that joint-liability incentives increased academic performance during classes both in the experimental courses and in the other simultaneous courses of the semester. Although this positive effect decreased during the period of exams, the overall impact of group incentives on academic performance was positive. To evaluate potential hypotheses regarding these observed patterns, at the conclusion of the study we conducted focus group-based discussions with students who had participated in the experiment. The explanations provided by these students about the fade-out of the effects of the intervention were consistent with the model proposed by Becker and Murphy, (1988); and later applied by Charness and Gneezy, (2009) in a field experiment about the formation of fitness habits. Specifically, based on the model, peer monitoring may increase human capital accumulation and facilitate the development of habit formation. This greater stock of human capital may have had positive effects on the academic performance of all the courses in the semester; however, joint-liability incentives may not have facilitated the development of 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 were absent. As articulated by Charness and Gneezy (p.927), "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 discount the future too much. As a result, people may underinvest in habit-forming activities."

To be effective, reinforced practice requires repeated trials or performance of the behavior followed by reinforcing consequences (Kazdin, 2012). Performing a behavior a few times may be not sufficient to develop a skill or habit. In the case of the current results, it is possible that completing homework during the period of classes contacted sufficient reinforcement to maintain the behavior but during the period of the final exams, there was not sufficient reinforcement to sustain responding. Thus, extinction (the removal of contingent reinforcement for a behavior that results in the extinguishing of the behavior; Kazdin, 2012) may have been in place during the final exam period resulting in a decrease in the behavior.

Also, it is possible that a conditioned reinforcer (e.g., attention, approval, affection from others—a generalized conditional reinforcer; see Kazdin, 2012) may have impacted the results. During the period of final exams, because students do not attend classes, not only were homework reinforcers absent but also potential conditioned reinforcers (e.g., student misses praise, smiles, and affection from the rest of the group).

An additional possible reason for our results was a potential peer effect. As the students in joint-liability group had superior performance on the midterm exams, this may have signaled to the students in the control group that they should (a) study more effectively for final exams, (b) obtain class-notes from the joint-liability students, and/or (c) study with the joint-liability students. Said another way, it is possible that the control group "caught up" with the joint-liability students as a result of interacting with them during the course. Also, it is possible that the use of extrinsic reinforcers and their subsequent removal might have negatively impacted the performance of the joint-liability students (e.g., Gneezy et al., 2011; Visaria et al., 2016).

Given previous findings that have shown positive effects of attendance on academic performance (i.e., Dobkin et al., 2010), it seems reasonable that the joint-liability incentives arrangement results might have been impacted by higher attendance rate of students because of pressure associated with the peer monitoring component. However, in the current study, attendance did not appear to be the cause of the superior performance of the students assigned to the joint-liability arrangement because they did not show a higher attendance rate.

Another potential explanation for our findings regarding the positive impact of joint incentives on take-home and midterm exams may involve a downward adjustment of effort. Tommasi and Weinschelbaum (2007) described a principal-agent relationship where the principal (e.g., the instructor) offers a contract to the agents (e.g., students) to elicit a high level of effort from the agents. The contract is designed to utilize peer monitoring as a potential mechanism of action. The agent can accept the contract (e.g., the student commits to exert considerable effort to accomplish the homework to receive an extra reward by the instructor) but unwind part of these incentives through

additional trades (e.g., when the student reaches a certain level of effort in that course, he actively rebalances his effort considering all of his courses, he trades effort between his courses, thus unwinding the pressure of the instructor to exert all the possible effort only in his course). Tommasi and Weinschelbaum referred to these outside trading opportunities as insurance. The main function of these potential trades is decreased risk for the agents, hence playing an insurance role. With regard to the current study, the students assigned to the joint-liability incentives condition were incentivized, as a result of the peer monitoring component, to increase the effort they devoted to the course. However, students in the current study took not only the experimental course, but also four or five other courses per semester. As such, they may have been desensitized from devoting a great deal of attention to a single course given they were enrolled in multiple courses.

In terms of our experiment, this mechanism would work in the following way. Students wish to demonstrate a satisfactory performance in their overall academic semester (i.e., in the 4–5 courses that they usually take per semester). Simultaneously, the instructor wishes to elicit a high level of effort from them in his course. Within a joint-liability arrangement, 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 (e.g., sit down and study together) which might alleviate a self-control problem. Thus, peer monitoring motivates some people to surpass a threshold needed to effectively engage in learning, at least for some time. When they have incurred this sunk cost, students might devote time to their classmates not only to study for the experimental course but also for the other simultaneous courses of the semester because they seek to achieve satisfactory performance in their overall academic semester. Thus, students assigned to the joint-liability group exhibited enhanced academic performances in homework and midterm examinations across all classes. Their positive academic experiences during semester across classes may have created a sense of self-efficacy because students were provided with evidence that they were capable of succeeding in the task (Dochy et al., 2011). At the time of final exams, when the peer monitoring element ceased, students might have dismissed the incentive to obtain better grades in final exams and relied on the higher grades obtained via homework and midterms during the semester; thus, decreasing their time devoted to studying for final exams. The extent to which the students could decrease their allocation of time and effort toward other classes might have been limited by the university requirement of a minimum grade of six (on a 1–12 scale) on the final examination to pass the course. Hence, the overall academic performance during the semester might have improved because each course is assessed via a combination of grades in homework and midterm exam (50 percent; when the peer monitoring element was present) and the grade in the final exam (50 percent; peer monitoring element was not present). In sum, the joint-liability incentive did not appear to harm performance in other simultaneous courses; rather the results suggested that they increased the students' overall academic performance.

A final implication of our research on a joint-liability system of incentives is that it could provide not only novel implications for the design of grading policies in the classroom but also for other social settings where incentives may be based on peer monitoring or joint liability. These ideas can be developed and tested in a

broader class of issues (e.g., Brown & Redmon, 1990; Dallery, et al., 2013; Kozica, 2020; McNally et al., 1983; Slavin et al., 1981), such as performance pay for teachers (e.g., where a joint-liability contract may be established, tying the earning of a bonus to the performance of other teacher´s students, and thus favoring peer monitoring between teachers); to design contracts between neighbors to reduce electric consumption in an apartment building; or between workers to increase output in a factory or office. In all of these settings, peers (e.g., other teachers, neighbors, colleagues) may 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.

Conclusions

Several conclusions may be drawn from this randomized field experiment. First, joint-liability incentives increased academic performance in the course and produced positive spillover effects on the other simultaneous courses during the semester attended by the students in the joint-liability incentives group. The main drawback of these positive effects of joint-liability incentives was the decrease in the rate of satisfaction toward their classmates reported by joint-liability students.

Second, the program appears to be very cost-effective. Specifically, the program entails an effective mechanism that does not include the use of tangible reinforcers. Third, individual incentives showed no effect on academic performance but did seem to negatively impact student evaluations of the instructor. Fourth, 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 pattern in the results including (a) the positive impacts of joint-liability incentives may diminish over time or (b) the control group may have caught up with the treatment group as a result of peer or signaling effects (i.e., because grades of classmates were known by all the students as grades were public, control group students could have identified that treatment students achieved better grades at homework and midterm exams and chose to study with them for the final exam). Further, students may have sought a satisfactory performance in all the courses and not focused solely on their grade just in the experimental courses. Specifically, students assigned to Treatment 1 Group (joint-liability incentives) were obliged, by means of peer monitoring, to increase the effort they devoted to the course. However, students took not only the experimental course, but also four or five other courses per semester. Thus, it is reasonable that they wanted to obtain a satisfactory overall performance and may not have been interested in devoting a great deal of attention to a single course. Thus, they may have decreased the amount time they devoted to studying for the final exam of the joint-liability course. Further research could further explore these possibilities.

Another question pertaining to joint-liability incentive systems refers to the potential effect of class size as well as group size on the efficiency and effectiveness of the arrangement. In the current study, the group size was small, which might have made it easier for students to monitor the behavior of classmates. This, in turn,

might have facilitated completion of homework by the monitored students. It is possible that in larger groups the effort associated with peer monitoring might result in less effective monitoring and subsequent less completion of homework by monitored students. It is possible that a larger group size might discourage a committed student in that there might be a greater probability that someone in the group will not fulfill the requirements to obtain the prize.

Future research should also evaluate potential differential long-term effects of joint-liability incentives across different students. For example, future research might assess what would happen if the additional incentive is reduced permanently (e.g., lower effort than before the extrinsic incentives were offered). Negative long-term effects on the reinforcing effects of students learning would be particularly troublesome (Gneezy et al., 2011). For example, incentives for exercise among undergraduate students have been shown to produce declines in exercising after the removal of incentives; particularly among those who already attended the gym regularly (Charness & Gneezy, 2009). Such potential effects on academic engagement and performance should be evaluated.

Future studies could also evaluate the relative and/or combined effects of peer monitoring and a potential "teamwork" effect. The joint-liability arrangement, which was originally designed to favor peer monitoring, may also increase cooperation between students and enhance performance. For instance, in the focus group, it was indicated that committed students in the joint-liability group encouraged those members who were demonstrating problems and poor academic performance to study together so as not to lose the bonus. Although working together in groups was not a requirement to earn the bonus, teamwork might have been a collateral effect of the joint-liability arrangement that might have impacted the results separate from the incentive element. Other future research might test whether the effects of joint incentives might be differentially impacted by how groups are formed (e.g., randomized vs. self-formed).

Finally, given the questionable efficacy of individual extrinsic incentives, educators may seek ways to make the learning experience more interesting. Said another way, if students develop an intrinsic motivation to improve their knowledge and skills, they may become fully engaged with learning and devote more effort to the 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). Future research should continue to explore this hypothesis.

Limitations

The external validity of our conclusions is limited in principle to students similar to those that participated in this field experiment. Despite this, it is important to consider that there was nothing atypical about the course characteristics, which were typical of first-year introductory courses in most universities. It is also unclear whether the results of the current study may be generalized to younger students.

Ideally, subsequent future investigations will study the extent to which the current results may be observed with young students.

Funding Not applicable.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Ethical Approval All the research was performed following ethical standards. Institutional Review Board (IRB) Approval: The project was approved by the Institutional Review Board at Universidad de Montevideo, resolution A 22-08-13: voluntary and informed participation; respect for the rights and dignity of participants, including confidentiality and anonymity; conducted with integrity and transparency; appropriateness of the local research environment and facilities.

References

Angrist, J., Oreopoulos, Ph., & Williams, T. (2014). When opportunity knocks, who answers? New evidence on college achievement awards. *Journal of Human Resources, 49*(3), 572–610.

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics, 1*(1), 136–163.

Banerjee, A., & Duflo, E. (2010). Giving credit where it is due. *Journal of Economic Perspectives, 24*(3), 61–79.

Becchetti, L., & Pisani, F. (2010). Microfinance, subsidies and local externalities. *Small Business Economics, 34*(3), 309–321.

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy, 96*, 675–700.

Bishop, J. (2006). Drinking from the fountain of knowledge: Student incentive to study and learn–externalities, information problems and peer pressure. *Handbook of the Economics of Education, 2*, 909–944.

Brown, C. C., Topping, K. J., Henington, C., & Skinner, C. H. (1999). Peer monitoring of learning behaviour: The case of 'checking chums'. *Educational Psychology in Practice, 15*(3), 174–182.

Brown, N., & Redmon, W. K. (1990). The effects of a group reinforcement contingency on staff use of unscheduled sick leave. *Journal of Organizational Behavior Management, 10*(2), 3–17.

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

Dal Bó, E., & Rossi, M. (2011). Term length and the effort of politicians. *Review of Economic Studies, 78*(4), 1237–1263.

Dallery, J., Raiff, B. R., & Grabinski, M. J. (2013). Internet-based contingency management to promote smoking cessation: A randomized controlled study. *Journal of Applied Behavior Analysis, 46*(4), 750–764.

Davies, S., & Witte, R. (2000). Self-management and peer-monitoring within a group contingency to decrease uncontrolled verbalizations of children with attention-deficit/hyperactivity disorder. *Psychology in the Schools, 37*(2), 135–147.

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

Dochy, F., Segers, M., & van Dinther, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review, 6*, 95–108.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *The American Economic Review, 101*(5), 1739–1774.

Fryer, R. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics, 126*, 1755–1798.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York city public schools. *Journal of Labor Economics, 31*(2), 373–407.
Fryer, R. G., Jr., Levitt, S. D., List, J., & Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. *National Bureau of Economic Research*. https://doi.org/10.3386/w18237
Ghatak, M., & Guinnane, T. W. (1999). The economics of lending with joint liability: Theory and practice. *Journal of Development Economics, 60*(1), 195–228.
Gneezy, U., Meier, S., & Rey-Biel, P. (2011). *Journal of Economic Perspectives, 25*(4), 1–21.
Grant, D., & Green, W. (2013). Grades as incentives. *Empirical Economics, 44*, 1563–1592.
Kazdin, A. E. (2012). *Behavior modification in applied settings*. Waveland Press
Kling, J., Liebman, J., & Katz, L. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83–119.
Kozica, J. (2020). Investigating the effects of self-monitoring, goal setting and a group contingency for increasing physical activity among adults with an intellectual disability (Doctoral dissertation, ResearchSpace@ Auckland
Larsen, Y. C., Groß, J., & Bogner, F. X. (2020). Bringing out-of-school learning into the classroom: Self- versus peer-monitoring of learning behaviour. *Education Sciences, 10*(10), 284.
McNally, R. J., Norusis, P. L., Gentz, S. A., & McConathy, L. C. (1983). Use of self-delivered reinforcement and group contingency management to increase productivity of severely retarded workers. *Psychological Reports, 52*(2), 499–503.
Montgomery, R., Bhattacharya, D., Hulme, D. (1996). Credit for the poor in Bangladesh: the BRAC Rural Development Programme and the Government Thana Resource Development and Employment Programme. In: Hulme and Mosley
Morrison, L., Kamps, D., Garcia, J., & Parker, D. (2001). Peer mediation and monitoring strategies to improve initiations and social skills for students with autism. *Journal of Positive Behavior Interventions, 3*(4), 237–250.
Pozo, S., & Stull, Ch. A. (2006). Requiring a math skills unit: Results of a randomized experiment. *American Economic Review, 96*(2), 437–441.
Price, M., Handley, K., Millar, J., & O'donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education, 35*(3), 277–289.
Rassuli, A. (2012). Engagement in classroom learning: Creating temporal participation incentives for extrinsically motivated students through bonus credits. *Journal of Education for Business, 87*(2), 86–93.
Slavin, R. E., Wodarski, J. S., & Blackburn, B. L. (1981). A group contingency for electricity conservation in master-metered apartments. *Journal of Applied Behavior Analysis, 14*(3), 357–363.
Smith, H. M., Evans-McCleon, T. N., Urbanski, B., & Justice, C. (2015). Check-in/check-out intervention with peer monitoring for a student with emotional-behavioral difficulties. *Journal of Counseling & Development, 93*(4), 451–459.
Swinton, O. H. (2010). The Effect of effort grading on learning. *Economics of Education Review, 29*, 1176–1182.
Speroni, C., Wellington, A., Burkander, P., Chiang, H., Herrmann, M., & Hallgren, K. (2020). Do educator performance incentives help students? Evidence from the teacher incentive fund national evaluation. *Journal of Labor Economics, 38*(3), 843–872.
Tommasi, M., & Weinschelbaum, F. (2007). Principal-agent contracts under the threat of insurance. *Journal of Institutional and Theoretical Economics, 163*(3), 379–393.
Visaria, S., Dehejia, R., Chao, M. M., & Mukhopadhyay, A. (2016). Unintended consequences of rewards for student attendance: Results from a field experiment in Indian classrooms. *Economics of Education Review, 54*, 173–184.
Zubrickas, R. (2013). Optimal grading. Available at SSRN-id2222645