International Review of Economics Education 27 (2018) 34–44

Contents lists available at ScienceDirect

International Review of Economics Education

journal homepage: www.elsevier.com/locate/iree

Monetary incentives and self-chosen goals in academic performance: An experimental study☆

Noemí Herranz-Zarzoso, Gerardo Sabater-Grande*

LEE & Economics Department, University Jaume I, Av. De Vicent Sos Baynat, s/n 12071, Castellón, Spain

ARTICLE INFO

JEL classification:
C93
D03
I21
J24

Keywords:
Self-chosen grades
Monetary incentives
Academic performance
Piece-rate mechanism
Rank-order tournament

ABSTRACT

This paper analyzes the effect of incentive-compatible self-chosen goals on academic performance by means of a randomized field experiment. We use two alternative payment mechanisms, a piece-rate and a rank-order tournament, to motivate students depending on their absolute or relative academic performance respectively. Students enrolled in Introductory Microeconomics were classified in two types depending on whether they had a failed background in this course (returning students) or they had not (new students). Controlling for potential confounding factors such as gender, degree, professor and university entrance grade, we find that both payment mechanisms are effective increasing grades of new and returning students.

1. Motivation

Many students are especially prone to focus too much on the present. Assessing the present costs of studying is much easier than evaluating (distant) potential future benefits. Policy makers might see this myopic behavior as an opportunity of improvement, and might want to try to offset present costs implementing closer benefits. One approach to address the present bias is simply requiring students to think about their academic goals and to formulate them. Another approach to correct the aforementioned bias is offering immediate incentives that trim immediate costs. Adopting both approaches, we conduct a randomized field experiment where over 170 undergraduate students are asked to report their individual goals on academic performance and monetary incentives are delivered to participants who reach their self-chosen goal.

2. Literature review

Our paper is related to two separate lines of research. The first one is the literature on goal setting. Goal setting is a cognitive theory based on the premise that the source of motivation is the desire and intention to reach a goal, i.e. the aim of a task that a person consciously desires to achieve or obtain (Locke and Latham, 2002; Locke and Latham, 2006). In achievement environments, such as higher education, task goal setting is a function of many variables, including skills and outcome expectations. Following Zimmerman (2011), outcome expectation can be defined as a belief about the success of a given task, differentiating it from the highly correlated concept of self-efficacy expectation, which is the belief about the personal capability to execute the behavior needed

☆ This work has been supported by the Ministerio de Economtria, Industria y Competitividad [grant ECO2015-68469-R]; and by the Ministerio de Educación, Cultura y Deporte [FPU014/02966] of the Spanish government

* Corresponding author.
E-mail addresses: nherranz@uji.es (N. Herranz-Zarzoso), sabater@uji.es (G. Sabater-Grande).

https://doi.org/10.1016/j.iree.2018.02.002
Received 5 September 2017; Received in revised form 21 November 2017; Accepted 26 February 2018
Available online 27 February 2018
1477-3880/

to produce the desired outcome. Both of them are key for students' motivation: if a student does not consider himself capable or has low outcome expectations, his motivation decreases and he does not make the effort needed to succeed.[1] Although Bandura (2006) developed a guide for constructing scales to measure perceived academic self-efficacy, it is not possible to get an incentive-compatible elicitation of self-efficacy scales. However, outcome expectations can be elicited using betting on outcomes as an incentive-compatible method. Specifically, when students bet on their outcome in a particular task, they take responsibility and ownership for their own goal. As Elliot and Fryer (2008) pointed out, such self-chosen goal setting is empowering and proactive, creating commitment and acceptance. In the same vein, Royer et al. (2015) and Samek (2016) concluded that giving the option to subjects to choose their own goal "acts as an internal commitment device meant to overcome problems of self-control" (Samek 2016, 2).

From the pioneering work of Latham and Locke (1979) to the present, there has been an extensive body of empirical literature testing goal-setting theory in controlled environments.[2,3] However, when it comes to self-chosen goals, the number of empirical contributions is limited and not all of them use incentive compatibility. Without financial incentives, Falk and Knell (2004) presented a social comparison model where people choose their own reference standards in order to accomplish goals of self-enhancement and self-improvement. The model's prediction about people tending to compare themselves to similar others was confirmed through a questionnaire where students only received a small show-up fee. Using the same methodology, Sackett et al. (2014) presented a questionnaire where marathoners were either asked or not asked to provide a time goal prior to their race. They found that the mere fact of asking runners (without potential financial reward) about their goals prior to the race improved performance among experienced but not novice marathoners.

Using a large sample of college students, Clark et al. (2016) analyzed the effect of self-chosen goals based on course performance and a specific task (completing online practice exams). They found that course goals had no significant impact on the performance of college students but, in contrast, task-based goals had a large positive effect on the level of task completion. However, students' extrinsic motivation in these two field experiments could be critically undermined because students were not financially rewarded when they met their self-chosen goals. The importance of financial incentives was analyzed in Goerg and Kube (2012). By means of a natural field experiment where workers were hired to re-organize a library, these authors presented the workers with an incentives' contract that combined self-chosen goals and monetary rewards in such a way that all of them preferred to set themselves a non-trivial goal. They found that self-imposed goals could work even in the absence of corresponding monetary incentives.

Empirical evidence on self-chosen goals where subjects' performance is based on financial incentives is even scarcer. Using a field experiment in an Indian data-entry firm, Kaur et al. (2010) tested whether workers demand self-disciplining devices. They found that a fraction of workers voluntarily agreed to incur in a monetary loss when falling short of a self-chosen production goal. Moreover, Dalton et al. (2015) proposed a simple model of self-chosen goals and tested its predictions in the laboratory, finding that only men confirmed their model's predictions: they exerted greater effort under the self-chosen goal contract system than under a piece-rate contract.

More related to our experiment and involving first-year university students, van Lent and Souverijn (2016) analyzed the effect of setting a goal and increasing its ambitiousness using mentor-student meetings. They found out that treatment group students performed better as compared to students in the control group. Nevertheless, students who were challenged to set a higher goal performed significantly worse than comparable students in the goal treatment. Contrary to van Lent and Souverijn (2016), we ask students to formulate their own goal and bet on it without any external influence, rewarding them financially. Theoretically, in our experiment, students' self-chosen goals must generate more effective incentives than those given by a third party because students set their goals based on their capabilities and knowledge, creating their own individual motivation.

Second, our paper is related to the literature based on financial incentives on academic performance. Experimental economists are convinced that higher incentives will lead to more effort and higher performance. However, psychologists claim that incentives improve performance in "algorithmic" or repetitive tasks, but they are less effective, or even counterproductive, in "heuristic" tasks requiring creativity, concentration, or intuition. Because learning has generally been classified as heuristic, extrinsic grade incentives may not be effective motivators. The argument behind this claim is that the use of incentives could crowd out intrinsic motivations that are important to produce the desired behavior.[4] Leaving aside ethical issues about the convenience of using financial incentives to improve students' grades, incentives have become object of interest of economists, policymakers, and researchers in the last years. In the literature, mixed evidence has been obtained using randomized field experiments and natural experiments (using databases from state programs) to analyze the effects of financial incentives on students' academic performance.[5]

3. Research design

All the evidence analyzed in the previous section is far from being conclusive because it has been generated in experiments or quasi-experiments with a wide array of incentive specifications, incentivized performances, target students and timing of performances and payments. The following is an elaboration of how such characteristics are related to the effectiveness of financial incentives on academic performance, comparing our study to the literature.

[1] See Pajares (2008) for a literature review on self-efficacy and regulation learning.

[2] These authors were the first to report evidence that goals lead to a better performance as compared to not setting goals.

[3] See Dykstra (2015) for a review on empirical evidence supporting goal setting as a tool to increase individuals' performance.

[4] Gneezy et al. (2011) state that a potential conflict arises between the extrinsic and the intrinsic effects of the incentives, especially in areas like education, contributions to public goods and forming habits, in the short run and in the long run.

[5] See Lavecchia et al. (2014) for a survey of effectiveness of financial incentives in education.

To begin with, we deal with the specification of the incentives provided. This specification includes two general categories: incentives for inputs and incentives for outputs. The former means anything (reading books, doing homework, attending school, etc.) that can contribute to learning under the student's control. The latter refers to student achievement, generally measured through test scores or class grades. Although Fryer (2011) concluded that output experiments demonstrated less-promising results than the input ones, we focus on output incentives given that traditional price theory predicts that they are socially optimal.[6] Output incentives include financial aid intended for payment of education expenses like tuition, fees and books, and freely available monetary rewards.[7] We prefer to use money as incentive because, as Croson (2005) pointed out, everyone values it and it is non-satiable (more is always better). Additionally, it is worth to note that our field experiment has been specifically designed for research purposes and it is not part of any state program dedicated to improve academic results. This fact allows us to manipulate independent variables, choosing timing incentives and payment mechanisms (piece-rate system and rank-order tournament) under controlled conditions.[8]

Second, we focus on the incentivized performance. In most cases, achievements required are referred to accomplishing a minimum composite score over an academic course, involving different subject matters and skills.[9] Given our interest in self-chosen goals, we offer incentives in a specific subject and over a limited time horizon, trying to facilitate students' thinking over their goals. With this objective in mind, we choose to incentivize the Introductory Microeconomics final exam like in Leuven et al. (2011). Similar to us, Levitt et al. (2011) incentivized a specific task (a standardized test), focusing on short-term effort in achievement.

Third, we address the issue of target students, differentiating (a) primary/middle education and higher education students, (b) higher and low ability students and (c) voluntary and non-voluntary students.[10] Regarding (a), we are interested in higher education students given that monetary incentives' provision to under-age subjects can be a problematic issue. This is so given that when participants are minors, experimentalists have to consider the ways children of different ages view the value of a payment, and to ensure that the amount and method are age-appropriate, and it does not present undue influence. With respect to (b), students are classified based on their scores in different tasks, which can be more or less related to the incentivized performance. Using scores in two programs of mathematics in Dutch secondary education, Leuven et al. (2010) sorted students in high and low ability types. These authors found that high-ability students had higher pass rates and collected significantly more credit points whenever they were assigned to (larger) reward groups. In contrast, low-ability students appeared to achieve less when they were assigned to the large reward group. These negative effects for less-able students were consistent with the work by Camerer and Hogarth (1999), who found that the performance threshold could result in a binding participation constraint at the bottom of the ability distribution, resulting in zero incentive effects for low-ability students. In the same line and following a tournament rule to the 30 best performing students, De Paola et al. (2012) obtained that financial rewards increased high ability (above the median high school grade) students' performance while the effect was null for low ability participants. Looking for the closest relationship between students' academic background and their incentivized task, we classify them in two types: returning students (if they had a failed background in Introductory Microeconomics) and new students (if they had not). Contrary to the aforementioned literature, where the effect of incentives is low or null for low ability students, we obtain that incentives are effective, increasing returning students' academic performance. Regarding (c), we randomly assign students to a treatment group (where incentives are offered) or to a control group (where no incentives are offered once they have explicitly declared their will to participate). Our design, requiring voluntary participation, ensures that subjects are really interested in being included in the experiment. In the terminology we use below, only subjects interested in the offered incentive will be considered participants. In the vast majority of the literature, experiments included as participants subjects who did not explicitly declare any interest in the offered incentive. Exceptions are Leuven et al. (2011), where students had to select themselves into different tournaments, Jackson (2010), Cha and Patel (2010), Scott-Clayton (2011) and Patel and Rudd (2012), where students had to enroll in the corresponding state program, and De Paola et al. (2012), where students were asked to fill a participation form.[11]

Lastly, we deal with another significant issue: time intervals. Specifically, we focus on the waiting time between the call and the starting of the task, and the waiting time between the accomplishment of the task and the reception of incentives (rewards). On one hand, most of the studies that reward incentives based on test performance announced the incentives well in advance of the test. On the other, studies that announced incentives immediately before the test distributed the payoffs with an appreciable delay. The evidence on such delayed rewards is mixed. In this sense, O'Neil et al. (1995) and O'Neil (1997) found that students' effort may be increased by financial rewards offered at the time of the test. In the same vein, Levitt et al. (2011) found that all motivating power of the incentives vanished when rewards were handed out with a delay. Therefore, shortening as much as possible both time intervals is a recommendable strategy in order to obtain a significant impact on student performance.[12] In our case, both waiting times are

[6] Using different award schemes for primary and middle school students, Fryer (2011) found that paying for performance on standardized tests had little or no effect on the outcomes for which students received financial incentives.

[7] Many colleges and universities offer financial incentives in the form of merit scholarships. However, these incentives are of a different nature to freely disposable money. See, for example, papers by Henry et al. (2004), Cornwell et al. (2005), Angrist and Lavy (2009), Angrist et al. (2009), Scott-Clayton (2011), Sjoquist and Winters (2012) and Castleman et al. (2014).

[8] The vast majority of the papers use only one payment mechanism, generally the piece-rate system.

[9] See, for example, papers by Fryer (2011), Bettinger (2012), Angrist et al. (2014), Barrow et al. (2016) and Castleman et al. (2014).

[10] See Lavecchia et al. (2014) for a literature classification according to education level.

[11] In our paper, incentive effects are disentangled from sorting effects by means of an experimental design that allows us to obtain individual data on academic performance with and without incentives in the same Introductory Microeconomic course. Contrary to Leuven et al. (2011), we find that the difference in performance between students can be attributed entirely to incentive effects.

[12] Braun et al. (2011) is an example of good results in which the incentive was announced immediately before the test and the reward was distributed immediately after the test.

reduced to a minimum: incentives are offered after the midterm exam, allowing a maximum one month time span between the announcement of the experiment and the day of the final exam (similar to Leuven et al., 2011).[13] After the publication of the definitive grades (two weeks after the exam), participants are immediately paid in cash.

Summing-up, our experimental design aims to maximize the efficacy of monetary incentives on academic performance in a specific task, offering monetary incentives (according to a piece-rate system or a competitive ranking) on self-chosen goals to volunteer higher education students. Based on this design strategy, we propose the following hypotheses:

Hypothesis 1 (*(H1)***)**. Monetary incentives based on self-chosen academic goals should increase the academic performance obtained by both new and returning students.

Hypothesis 2 (*(H2)***)**. A piece-rate payment mechanism should be more effective upgrading average academic performance of all participants than a rank-order tournament system. Contrary to piece-rate payment schemes, under rank-order tournaments students' payoffs are discontinuous in the level of exerted effort. This is so given that a marginal unit of effort increases the expected student payoff by increasing the probability to win, but it does not necessarily generate a higher payoff. This feature can produce complicated behavioral effects affecting its effectiveness.

Our results confirm H1 and reject H2 using a clean experimental design where confounding factors such as gender, degree, professor and university entrance grades are controlled.

Onwards, the structure of the paper is organized as follows: first, we present the design of the implemented experiment; after that, we analyze the empirical evidence collected; and lastly, we present our conclusions.

4. Experimental design

We conducted a randomized field experiment aiming to improve academic performance in an introductory course of Microeconomics through monetary incentives offered on the basis of self-chosen goals.

New and returning students enrolled in the 2017/2018 course of Introductory Microeconomics at the University Jaume I were offered the possibility of taking part in a monetary incentive program asking them through an invitation call about their willingness to participate. We opened our call once students had been midterm examined.[14] In the call they were informed that students responding affirmatively would be randomly assigned[15] to one of three groups: a control group, Treatment 1 (T1), or two alternative treatment groups: Treatment 2 (T2) and Treatment 3 (T3). In T1 participants would not be monetarily incentivized. Alternatively, participants assigned to T2 would be paid according to their absolute (piece-rate) academic performance. Lastly, in T3 a rank-order tournament would be used as payment mechanism to reward participants.[16] Additionally, we notified that participants would receive information about their corresponding group before they were invited to choose a goal for their final exam grade. This design discards possible willingness effects given that, both the treatment groups and the control group are integrated only by students declaring their willingness to participate in the program. Moreover, students not responding our call were included in Treatment 0 (T0) in order to compare non-participants' midterm grades with the corresponding ones obtained by participants. In doing so, we can check for the actual existence of a (potential) willingness effect in our sample.

Like in Clark et al. (2016), we allowed participants to bet[17] according to their personal goal (as opposed to goals set by a professor).[18] In this way, the goal was tailored to each student's degree of self-control and ambition. In order to do so, we used an experimental design strategy similar to Jackson (2010), opening a call named "Bet for your grade and win", right after students had been informed about their grades in a midterm exam.[19] This within-subject strategy aims to collect individual grades with and without incentives for the same subject. Among 496 students enrolled in Introductory Microeconomics (406 new students and 90 returning students), 177 of them (111 new students and 66 returning students) attended the call to participate in this experiment. These rates of acceptation contrast with De Paola et al. (2012) where about 90% of students assigned to treated groups decided to participate in the experiment.

Table 1 lists the summary of the treatments discussed above.

In the experiment, the purpose of bettors was to maximize their monetary reward (R). Subjects were informed that their R would depend on their bet (B), the grade (G) they obtained in the Introductory Microeconomics final exam, and, only for returning students, the average grade (AG) in the same subject-matter final exams in previous semesters, according to the following function:

[13] Students were allowed to bet until the day before the final exam.

[14] Subjects had not been informed about any incentive program before the midterm exam.

[15] We randomize the treatments within each type (new and returning) of student.

[16] Rank-order tournaments as analyzed in Lazear and Rosen (2009) are characterized by the evaluation of individual performance relative to the performance of competitors. Regarding this issue van Dijk et al. (2001) find that tournaments lead to a higher effort on average but more variable compared to the other payment schemes.

[17] Note that we use the meaning of "bet" as a guess or opinion, given that our participants do not risk their own money.

[18] Unlike van Lent and Souverijn (2016), where a mentor-student meeting was used to induce students to set a course-specific grade goal, we choose to implement a website to gather students' bets in order to avoid any kind of bias in their outcome expectations.

[19] In order to participate in the experiment, subjects were asked to register on a betting system based on PHP + MySQL through our lab's (LEE) website. Once registered they were randomly assigned to the control or the treatment groups. Then, all students could bet on the highest grade they thought they could get.

Table 1
Summary of treatments. ME: Midterm exam; FE: Final Exam.

Treatments	*Number of subjects*		*Treatment variables*		
	New	*Returning*	*Willingness to participate*	*Incentive ME/FE*	*Payment Mechanism*
T0	295	24	No	No/No	–
T1	41	19	Yes	No/No	–
T2	31	22	Yes	No/Yes	Piece-rate
T3	39	25	Yes	No/Yes	Tournament

$$R = \left(B - \left[\frac{AG}{2}\right]\right)^2 \; \forall \; G \geq B$$

Students were rewarded with R euros only if G was higher than or equal to B.[20]

Therefore, in T2 all students whose grade was higher than or equal to their bet (successful bettors) earned R euros.[21] Alternatively, in T3 we implemented two rank-order identical tournaments, one for the new students and another one for returning students, offering the aforementioned rewards only for the top three students in each tournament.[22] Thus, new and returning students compete separately in two independent tournaments for prizes. In the experiment instructions, participants[23] were informed that the two rankings would be published soon after the notification of the final exam actual grades, and prizes would be delivered to participants immediately after that.[24]

Our experimental design allows us: (a) to make between-subjects comparisons between grades obtained in the midterm exam by non-participants (T0) and participants (T1 + T2 + T3) in order to test for willingness effects, (b) to make between-subjects comparisons between grades obtained in the final exam by T1 students and T2/T3 students and (c) to analyze the factors driving the incentivized grades and bets, controlling for potential confounding factors, such as gender, degree, professor and the University Entrance Grade (UEG).

5. Data analysis

5.1. Sample self-selection: a potential willingness effect

Because of our design requires students' willingness to participate in the program in both the control group and the treatments group, potential self-selection problems are ruled out. In spite of this, it can be interesting to analyze the real existence of this potential effect comparing grades obtained in the midterm exam between students who are willing and who are not willing to participate in the program.

Table 2 shows that the average grades obtained in the midterm exam by new/returning students declaring their willingness to participate in the program are higher than non-participants. However, these differences are statistical significant in median only for new students.[25] In consequence, requiring voluntary participation in the control group avoids an actual self-selection problem in our sample of new students.

Additionally, as double-check tool for rule out self-selection problems, we compare grades obtained by students in the midterm exam between our treatment groups (T2/T3) and the control group (T1) finding no significant differences in any case.[26]

5.2. Descriptive statistics

This section presents descriptive statistics of students' bets and grades corresponding to the final exam in the Introductory Microeconomics course.

5.2.1. Bets

Fig. 1 shows new and returning students' bets averages. For both, new and returning students, we observe that the bets' average is slightly higher in T3 compared to T2.[27] Moreover, the bets' averages obtained by new students are higher than the ones obtained by

[20] For new students $R = B^2$ given that $AG = 0$.

[21] Note that, if G ≥ B, the reward was the same for all students with identical bets, independently of the obtained grade. Additionally, students were only paid if their final grade was at least 5 out of 10.

[22] According to Vandegrift et al. (2007), a tournament in which second and third-place performers also receive a payment should induce lower performance than a winner-take-all tournament.

[23] New and returning students were informed about the number of participants included in their group before they were allowed to bet.

[24] The instructions are available upon request.

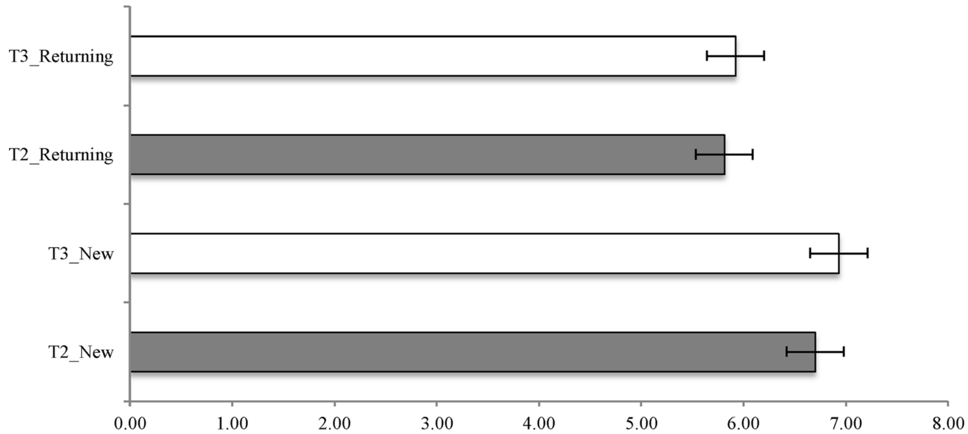
[25] Mann-Whitney test *p*-values for new (returning) students: 0.013 (0.291).

[26] Using a t- (Mann-Whitney) test we find no significant differences between T1 and T2, and T1 and T3, for both new (returning) students, *p*-values: 0.771 and 0.760 (0.236 and 0.398) respectively.

[27] Using a Mann-Whitney test we find no significant differences between T2 and T3 bets for both new (*p*-value: 0.935) and returning students (*p*-value: 0.270)

Table 2
Averages and standard deviations (included between parentheses) corresponding to grades obtained in the midterm exam by new and returning students.

Students	New		Returning	
Participants	Yes	No	Yes	No
	4.35 (2.72)	3.59 (2.58)	4.28 (2.45)	3.78 (2.28)



Students	New		Returning	
Treatment	*T2*	*T3*	*T2*	*T3*
	6.70 (1.10)	6.93 (1.35)	5.81 (0.83)	5.92 (0.83)

Fig. 1. Bets' average (standard deviations included between parentheses).

returning students.

In view of the fact that only the top three students were monetary rewarded in T3, bettors in this treatment could be compelled to overestimate their goals, resulting in higher bets.

5.2.2. Grades

For returning students, we define "grade improvement" as the students' final exam grade in the semester in which the incentive program was implemented minus the average of final exam grades in previous semesters.

Fig. 2 displays descriptive statistics corresponding to new students' grades and returning students' grade improvements after our call. We observe that the average of grades obtained by T2 (T3) new students is 1.36 (1.28) points higher than the average of grades obtained by T1 students. A similar pattern is observed for returning students given that the difference between T2 (T3) and T1 average of incentivized improvement grades is 1.23 (1.43). Graphically, we can see these aforementioned differences in grades and grade improvements between the control group and the treatment group students for new and returning students respectively are noticeable.

Additionally, for new students in the two treatment groups, we observe that the standard deviation of incentivized grades is higher than the one corresponding to T1. Nevertheless, for returning students we observe the opposite effect.

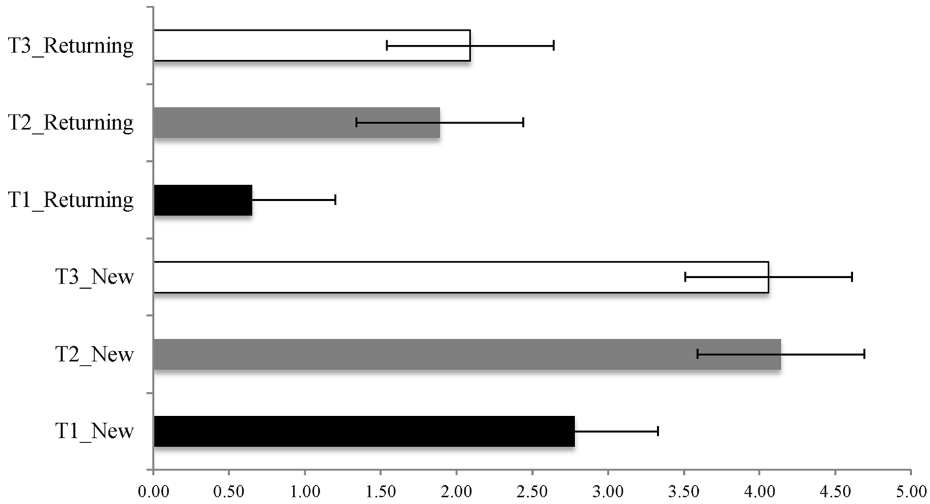
Comparing the treatment groups, T2 and T3, we observe no noticeable differences between grades under the two incentive systems (piece-rate and rank-order tournament) for both new and returning students.

5.3. Statistical tests

5.3.1. Incentive effects

In this subsection, we analyze the effect of monetary incentives on new and returning students' final exam grades. Since all datasets are normally distributed for a 95% level of confidence, parametric tests are used in order to compare incentivized grades: (a) a *t*-test comparing means from the two populations, (b) a Levene test comparing variances and (c) a Kolmogorov-Smirnov test comparing distributions.

5.3.1.1. New students. Distributions of grades obtained in the final exam by students in T1, T2 and T3 are presented in Fig. 3. For



Students	New			Returning		
Treatment	*T1*	*T2*	*T3*	*T1*	*T2*	*T3*
	2.78 (1.65)	4.14 (2.96)	4.06 (2.63)	+0.65 (2.30)	+1.89 (2.20)	+2.09 (1.62)

Fig. 2. Averages and standard deviations (included between parentheses) corresponding to grades (new students) and grade improvements (returning students).

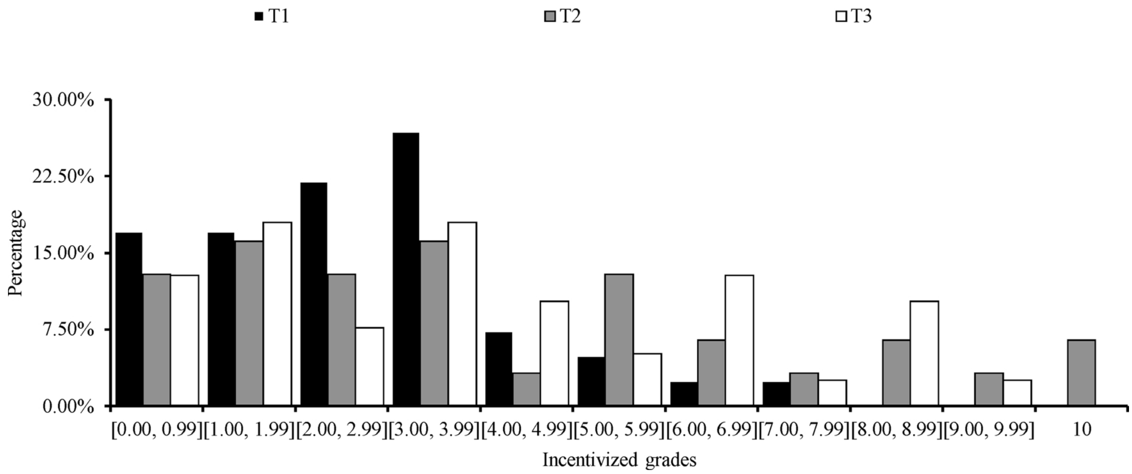


Fig. 3. Distribution of grades obtained in the final exam by new students in T1, T2 and T3.

grades higher than five, the percentage of treatment group students is higher than the controls' percentage. Particularly, only incentivized students reach the highest grades. Additionally, for a 90% level of confidence, grades obtained by new students in T2 can be considered not normally distributed.[28] For this reason, we implement additional non-parametric tests in order to confirm our results.

For both treatments, T2 and T3, from (a) we find that the median of grades obtained by incentivized groups is significantly higher than the median of grades obtained by students in our control group.[29] In addition, we obtain from b) that the variance of T2/T3 grades is significantly higher than the variance of T1 grades,[30] while from (c) we find that there are no significant differences between the distributions of grades obtained by T2/T3 students and the corresponding distribution for T1 students.[31]

Result 1: For both payment schemes, piece-rate and rank-order tournament, monetary incentives based on self-chosen goals are

[28] Kolmogorov-Smirnov test *p*-values for T1, T2 and T3 are 0.084, 0.119 and 0.115.

[29] T-test *p*-values: 0.027 for T2 and 0.012 for T3. Additionally, Mann-Whitney test *p*-values are 0.096 for T2 and 0.042 for T3.

[30] Levene test *p*-values: 0.000 for T2 and 0.002 for T3.

[31] Kolmogorov-Smirnov test *p*-values: 0.104 for T2 and 0.079 for T3.

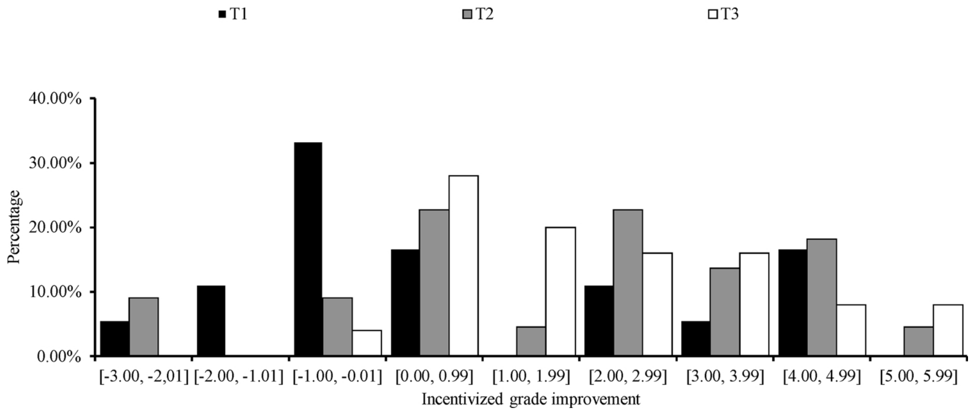


Fig. 4. Distribution of grade improvements obtained by returning students in T1, T2 and T3.

effective to increase grades obtained by voluntary new students

5.3.2. Returning students

Fig. 4 presents distributions of improvement grades obtained in the final exam by returning students in T1, T2 and T3. In this figure, we can observe that, with the exception of improvement grades between 4 and 4.99, for any positive grade improvement the percentage of controls is lower than the percentage of incentivized students.

In both treatments, from (a) we find that mean of grade improvements in T2/T3 is significantly higher than the mean of grade improvements obtained by non-incentivized students.[32] Moreover, from (b) we find significant differences between the variances of grade improvements only comparing T1 and T3,[33] while from c) significant differences are obtained between distributions of incentivized and control students.[34]

Result 2: Both relative and absolute academic monetary incentives based on self-chosen goals are effective to increase grade improvements of returning voluntary students.

Results 1 and 2 confirm our H1. However, result 2 contrasts with Grove and Wasserman (2006) that using a natural experiment concluded that grade incentives to practice economics throughout the semester boosted the average freshman exam performance, but not that of academically above- or below- average students, or of any other category of students.

5.3.3. The effects of different payment mechanisms

In this section, we test H2 comparing the treatments groups (T2 and T3) in order to analyze the effect of the two payment mechanisms implemented for both new and returning students. Contrary to our H2, we find that there are no significant differences between medians, variances and distributions of (improvement) grades obtained by new (returning) students in T2 and the corresponding ones in T3.[35]

In accordance with the findings obtained by van Dijk et al. (2001), that workers with relatively low ability do not seem to realize that they have little chance of winning a rank-order tournament, we obtain that subjects do not bet differently owing to the payment system. Namely, differences in the median/distribution of bets are not statistically significant between treatments.[36]

Result 3: Incentives based on relative academic performance are equally effective than those based on absolute performance to increase the average of students' grades

5.4. Regression analysis

5.4.1. Determinants of bets

In this subsection, we estimate different models to shed light on the determinants of the bets.

The models include as covariates: (1) the fact of being assigned randomly to T1 (where students are paid according to a piece rate) or to T2 (where a tournament is used to reward students), (2) the non-incentivized (midterm) grades collected before the starting of the incentives' program, (3) the fact of being a returning student, (4) the degree they are enrolled in,[37] (5) professor[38] and (6) the UEG.

[32] T-test *p*-values: 0.080 for T2 and 0.027 for T3.

[33] Levene test *p*-values: 0.708 for T2 and 0.088 for T3.

[34] Kolmogorov-Smirnov test *p*-values: 0.066 for T2 and 0.011 for T3.

[35] For new (returning) students, t, Levene and Kolmogorov-Smirnov tests *p*-values are 0.903 (0.716), 0.458 (0.092) and 0.982 (0.357) respectively.

[36] Mann-Whitney test *p*-value is 0.761 and Kolmogorov-Smirnov test *p*-value is 0.249.

[37] The subject Introductory Microeconomics is shared by students of 4 different degrees: Management, Economics, double degree of Economics and Law and Accounting.

[38] Each professor teaches one of the 6 class groups from A to F.

Table 3
Models explaining bets.

Bet	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Non-incentivized grade	0.0906**	0.0908**	0.0725*	0.0842**	0.0638	0.0450	0.0533
	(0.0424)	(0.0408)	(0.0402)	(0.0397)	(0.0394)	(0.0437)	(0.0436)
Returning		−0.722***	−0.984***	−0.811***	−0.647**	−0.618**	−0.824**
		(0.254)	(0.262)	(0.268)	(0.268)	(0.301)	(0.325)
T2			0.807**	0.804**	0.628*	0.635*	0.621*
			(0.331)	(0.324)	(0.326)	(0.325)	(0.322)
T3			0.907***	0.914***	0.883***	0.926***	0.925***
			(0.321)	(0.314)	(0.307)	(0.320)	(0.317)
Gender				0.524**	0.582**	0.560**	0.491*
				(0.238)	(0.245)	(0.262)	(0.264)
Management					−0.173	−0.237	−0.264
					(0.279)	(0.282)	(0.280)
Economics					−0.214	−0.503	−0.451
					(0.304)	(0.330)	(0.329)
Economics + Law					1.219**	1.290**	1.831***
					(0.492)	(0.534)	(0.628)
Class group A						−0.0331	0.207
						(0.329)	(0.359)
Class group B						−0.0858	−0.0377
						(0.405)	(0.402)
Class group D						0.233	0.268
						(0.487)	(0.483)
Class group E						0.122	0.151
						(0.386)	(0.383)
Class group F						1.090**	1.000**
						(0.478)	(0.477)
UEG							−0.147
							(0.0923)
Constant	5.882***	6.104***	5.602***	5.185***	5.331***	5.398***	6.575***
	(0.248)	(0.251)	(0.298)	(0.348)	(0.415)	(0.419)	(0.847)
Observations	94	94	94	94	94	94	94
R-squared	0.047	0.125	0.201	0.243	0.314	0.369	0.389

Standard errors in parentheses.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

The last OLS model presented (Model 7) in Table 3 controls for all the potential confounding factors. The results obtained in this analysis are quite conclusive. The midterm exam grades do not influence subjects' bets. Otherwise, other factors such as treatment group, the fact of being returning, gender, degree and class group affect significantly the self-chosen goals of students. Particularly, those who have failed the subject previously choose less demanding goals than new students. Furthermore, given that in the tournament mechanism (T3) only the top three students are rewarded, their bets are higher than the ones of those betting under a piece rate mechanism (T2). In the explanation of bets, gender plays a role, pointing that women are more conservative in terms of bets than men are. In addition, students enrolled in the double degree bet significantly higher than those enrolled in accounting do. A possible explanation may rest on the fact that they are, theoretically, cleverer given that the cut-off mark for entering in this degree is higher. Furthermore, one class group is statistically significant respect to class group C. We use class group C as a reference category because the professor is one of the experimentalists. The results help us to avoid potential critics about the problem of having a professor in the subject who has design the experiment given that only the bets fixed by students belonging to the class group F are significantly higher. The remaining class groups bet in a similar way. Finally, the UEG does not affect students' self-chosen goals.

5.4.2. Determinants of the incentivized grades

Now, we estimate different OLS models to shed light on the key determinants of the incentives' program effects on students' incentivized grades. This analysis not only supports our previous results based on descriptive analysis, but also controls for the potential influences of confounding factors. In Table 4 we include the same covariates as before.

As long as more covariates are included in the models the explanation power increases and, step-by-step the effect of each explanatory variable can be analyzed in a detailed way. The first result obtained is the power of the incentives provided increasing grades. Specifically, the tournament is more powerful increasing grades than the piece rate, although both of them reach their main purpose. The career path of the student in the same subject (non-incentivized grade) as a predictor of their grade is verified. Another determinant explaining incentivized grades is the fact of being a returning student in the subject: given that, it is not the first time they face the subject's concepts, their grade increases in average more than the one obtained by subjects enrolled in the subject the first time. Gender does not influence the grades under incentives. Additionally, another determinant is the degree they are enrolled in, pointing that those coursing the double degree in economics and law obtained significantly higher grades than those enrolled in

Table 4
Models explaining incentivized grades.

Incentivized grade	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
T2	1.354***	1.252***	1.146**	1.146**	0.928**	0.891**	0.863**
	(0.431)	(0.440)	(0.442)	(0.443)	(0.406)	(0.365)	(0.372)
T3	1.357***	1.395***	1.303***	1.303***	1.334***	1.141***	1.115***
	(0.414)	(0.419)	(0.420)	(0.421)	(0.384)	(0.348)	(0.354)
Non-incentivized grade		0.196***	0.199***	0.199***	0.136**	0.257***	0.262***
		(0.0669)	(0.0665)	(0.0668)	(0.0620)	(0.0606)	(0.0627)
Returning			0.626*	0.626*	1.065***	0.868**	0.839**
			(0.367)	(0.377)	(0.353)	(0.341)	(0.370)
Gender				0.00143	0.221	−0.240	−0.253
				(0.360)	(0.341)	(0.314)	(0.319)
Management					−0.394	−0.156	−0.153
					(0.416)	(0.380)	(0.385)
Economics					−0.123	0.135	0.130
					(0.463)	(0.428)	(0.432)
Economics + Law					3.952***	3.849***	3.983***
					(0.773)	(0.740)	(0.828)
Class group A						0.0634	0.123
						(0.445)	(0.487)
Class group B						2.177***	2.187***
						(0.470)	(0.475)
Class group D						−0.136	−0.142
						(0.553)	(0.558)
Class group D						−1.430***	−1.389**
						(0.525)	(0.540)
Class group F						−0.368	−0.393
						(0.573)	(0.580)
UEG							−0.0438
							(0.123)
Constant	2.955***	2.181***	1.999***	1.998***	2.043***	1.600***	1.952*
	(0.298)	(0.405)	(0.416)	(0.479)	(0.559)	(0.531)	(1.088)
Observations	177	158	158	158	158	158	156
R-squared	0.073	0.134	0.150	0.150	0.310	0.472	0.469

Standard errors in parentheses.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

accounting. The last factors explaining the grades obtained in the incentivized scenario are the class groups they belong to or in other words, the professor. As we have explained previously we use class group C as a reference category because the professor is one of the experimentalists. The grades obtained by his students are in the mean, finding groups with significantly more and fewer points in the incentivized grades analyzed thus, our results are not influenced by this fact. Lastly, the UEG has not any effect in the incentivized grades obtained by the students.

The previous results based on descriptive analysis are corroborated by means of these regressions making our results about the role of the incentives' program designed more convincing.

Result 5: Controlling for potential confounding factors both payment mechanisms are effective increasing grades of new and returning students. Furthermore, although grades obtained in the midterm exam do not affect students' self-chosen goals, they do explain their final exam grades.

6. Conclusions

Our paper is the first study that introduces self-chosen academic goals in an incentive-compatible field experiment including two alternative payment mechanisms. We have analyzed the effectiveness of monetary incentives based on absolute and relative academic performance using a piece-rate and a rank-order tournament system respectively. New and returning students in Introductory Microeconomics at the University Jaume I were offered the possibility to participate in the incentives program explaining them beforehand the different groups they could be assigned to. Those accepting and showing explicitly their will to participate were randomly assigned to a control (where no incentives were offered) or to two alternative treatment groups (where incentives were offered) and then, they were asked to bet on their own final exam grade. Our experimental design aims to maximize the efficiency of monetary incentives based on self-chosen goals, choosing to a specific task and a limited time horizon, and providing the incentives with immediacy.

Our results suggest that non-incentivized grades obtained in the midterm exam do not influence students' self-chosen goals, but they explain the incentivized grades obtained in the final exam.

Controlling for potential confounding factors as gender, degree, professor and the UEG, we find that incentives based on a tournament mechanism are as effective as those based a piece-rate scheme to increase grades of new and returning students.

Although this paper highlights the importance of providing incentives to students in order to increase their academic performance, policymakers could be worried by the financial resources needed to guarantee its effectiveness. Our results show that a (low-cost) rank-order tournament payment mechanism based on self-chosen goals can be as effective as a (resource intensive) piece-rate mechanism. Consequently, government policies should be able to support the development of effective incentive programs, allowing students to compete for rewards based on their self-chosen goals.

References

Angrist, Joshua, Lavy, Victor, 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. Am. Econ. Rev. 99 (4), 1384–1414.
Angrist, Joshua, Lang, Daniel, Oreopoulos, Philip, 2009. Incentives and services for college achievement: evidence from a randomized trial. J. Appl. Econ. 1 (1), 136–163.
Angrist, Joshua, Oreopoulos, Philip, Williams, Tyler, 2014. When opportunity knocks, who answers? New evidence on collage achievement awards. J. Hum. Resour. 49 (3), 572–610.
Bandura, Albert, 2006. Guide for constructing self-efficacy scales. In: Pajares, Frank, Urdan, Timothy C. (Eds.), Self-Efficacy Beliefs of Adolescents. Information Age Pub, Greenwich, CT, pp. 307–337.
Barrow, Lisa, Elena Rouse, Cecilia, 2016. Financial incentives and educational investment: the impact of performance-based scholarships on student time use. Educ. Finance Policy 1–55.
Bettinger, Eric P.., 2012. Paying to learn: the effect of financial incentives on elementary school test scores. Rev. Econ. Stat. 94 (3), 686–698.
Braun, Henry, Kirsch, Irwin, Yamamoto, Kentaro, 2011. An experimental study of the effects of monetary incentives on performance on the 12th-Grade NAEP reading assessment. Teach. Coll. Rec. 113 (11), 2309–2344.
Camerer, Colin F., Hogarth, Robin M., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. J. Risk Uncertain. 19 (1–3), 7–42.
Castleman, Benjamin L., Terry Long, Bridget, Avery, Chris, Murnane, Richard, Goodman, Josh, Castleman, L Benjamin, 2014. The Impact of Partial and Full Merit Scholarships on College Entry and Success: Evidence from the Florida Bright Futures Scholarship Program. pp. 17.
Cha, Paulette, Patel, Reshma, 2010. Rewarding Progress, Reducing Debt: Early Results from the Performance-Based Scholarship Demonstration in Ohio.
Clark, Damon, Gill, David, Prowse, Victoria L., Rush, Mark, 2016. Using Goals to Motivate College Students: Theory and Evidence from Field Experiments.
Cornwell, Christopher M., Hee Lee, Kyung, Mustard, David B., 2005. Student responses to merit scholarship retention rules. J. Hum. Resour. XL (4), 895–917.
Croson, Rachel, 2005. The method of experimental economics. Int. Negot. 10 (1), 131–148.
Dalton, Patricio, Gonzalez Jimenez, Victor, Noussair, Charles, 2015. Paying with Self-Chosen Goals: Incentives and Gender Differences.
De Paola, Maria, Scoppa, Vincenzo, Nisticó, Rosanna, 2012. Monetary incentives and student achievement in a depressed labor market: results from a randomized experiment. J. Hum. Cap. 6 (1), 56–85.
Dykstra, Angela, 2015. A Literature Review of Expectancy & Goal-Setting Theories When Meeting Organizational Goals and Effecting Change for Diverse Populations. Southern New Hampshire University.
Falk, Armin, Knell, Markus, 2004. Choosing the Joneses: endogenous goals and reference standards. Scand. J. Econ. 106 (3), 417–435.
Fryer, Roland G., 2011. Financial incentives and student achievement: evidence from randomized trials. Q. J. Econ. 126 (4), 1755–1798.
Fryer, James W., Elliot, Andrew J., 2008. Self-regulation of achievement goal pursuit, motivation and self-regulated learning: Theory, research, and applications. pp. 53–75.
Gneezy, Uri, Meier, Stephan, Rey-Biel, Pedro, 2011. When and why incentives (don't) work to modify behavior. J. Econ. Perspect. 25 (4), 191–210.
Goerg, Sebastian J., Kube, Sebastian, 2012. Goals (Th)at Work—Goals, Monetary Incentives, and Workers' Performance. MPI Collective Goods Preprint 2012/19.
Grove, Wayne A, Wasserman, Tim, 2006. Incentives and student learning: a natural experiment with economics problem sets. Am. Econ. Rev. 96 (2), 447–452.
Henry, G.T., Rubenstein, Ross, Bugler, Daniel T., 2004. Is HOPE enough? Impacts of receiving and losing merit-based financial aid. Educ. Policy 18 (5), 686–709.
Jackson, C. Kirabo, 2010. A little now for a lot later: a look at a texas advanced placement incentive program. J. Hum. Resour. 45 (3), 591–639.
Kaur, Supreet, Kremer, Michael, Mullainathan, Sendhil, 2010. Self-control and the development of work arrangements. Am. Econ. Rev.: Pap. Proc. 100 (2), 624–628.
Latham, Gary P., Locke, Edwin A., 1979. Goal setting—a motivational technique that works. Organ. Dyn. 8 (2), 68–80.
Lavecchia, Adam, Liu, Heidi, Oreopoulos, Philip, 2014. Behavioral Economics of Education: Progress and Possibilities. http://dx.doi.org/10.3386/w20609. 20609.
Lazear, Edward P., Rosen, Sherwin, 2009. Rank and order tournament: an optimal labor contract. J. Polit. Econ. 89 (5), 841–864.
Leuven, Edwin, Oosterbeek, Hessel, van der Klaauw, Bas, 2010. The effect of financial rewards on students' achievement: evidence from a randomized experiment. J. Eur. Econ. Assoc.
Leuven, Edwin, Oosterbeek, Hessel, Sonnemans, Joep, van der Klaauw, Bas, 2011. Incentives versus sorting in tournaments: evidence from a field experiment. J. Labor Econ. 29 (3), 637–658.
Levitt, Steven D, List, John A, Neckermann, Susanne, Sadoff, Sally, 2011. The Impact of Short-Term Incentives on Student Performance. Becker Firedman Institute for Research in Economics, University of Chicago.
Locke, Edwin A., Latham, Gary P., 2002. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am. Psychol. 57 (9), 705–717.
Locke, Edwin A., Latham, Gary P., 2006. New directions in goal-setting theory. Curr. Direct. Psychol. Sci. 15 (5), 265–268.
O'Neil Jr., Harold F., Sugrue, Brenda, Baker, Eva L., 1995. Effects of motivational interventions on the national assessment of educational progress mathematics performance. Educ. Assess. 3 (2), 135–157.
O'Neil, Harold F., 1997. Final Report of Experimental Studies on Motivation and NAEP Test Performance. Los Angeles.
Pajares, Frank, 2008. Motivation and self-regulated learning: beliefs in self-Regulated learning. In: Schunk, Dale H., Zimmerman, Barry J. (Eds.), Motivation and Self-Regulated Learning: Theory, Research, and Applications. Lawrence Erlbaum Associates, pp. 111–139.
Patel, Reshma, Rudd, Timothy, 2012. Can Scholarships Alone Help Students Succeed?.
Royer, Heather, Stehr, Mark, Sydnor, Justin, 2015. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. Am. Econ. J.: Appl. Econ. 7 (3), 51–84.
Sackett, Aaron M., Wu, George, White, Rebecca J., Markle, Alex, 2014. Harnessing optimism: how eliciting goals improves performance. SSRN Electron. J.
Samek, Anya Savikhin, 2016. Gifts and Goals: Behavioral Nudges to Improve Child Food Choice at School.
Scott-Clayton, Judith, 2011. On money and motivation: a quasi experimental analysis of financial incentives for college achievement. J. Hum. Resour. 46 (3), 614–646.
Sjoquist, David L., Winters, John V., 2012. Building the stock of college-educated labor revisited. J. Hum. Resour. 47 (1), 270–285.
Vandegrift, Donald, Yavas, Abdullah, Brown, Paul M., 2007. Incentive effects and overcrowding in tournaments: an experimental analysis. Exp. Econ. 10 (4), 345–368.
van Dijk, Frans, Sonnemans, Joep, Van Winden, Frans, 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45 (2), 187–214.
van Lent, Max, Souverijn, Michiel, 2016. Goal Setting and Raising the Bar: A Field Experiment. 2017/-001/VII.
Zimmerman, Barry J., 2011. Motivational sources and outcomes of self-regulated learning and performance. In: Zimmerman, Barry J., Schunk, Dale H. (Eds.), Handbook of Self-Regulation of Learning and Performance. Routledge, New York, pp. 49–64.