ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Bigoni, Maria; Fort, Margherita; Nardotto, Mattia; Reggiani, Tommaso

Working Paper

Teams or tournaments? A field experiment on cooperation and competition among university students

IZA Discussion Papers, No. 5844

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Bigoni, Maria; Fort, Margherita; Nardotto, Mattia; Reggiani, Tommaso (2011) : Teams or tournaments? A field experiment on cooperation and competition among university students, IZA Discussion Papers, No. 5844, Institute for the Study of Labor (IZA), Bonn, https://nbn-resolving.de/urn:nbn:de:101:1-201107283696

This Version is available at:
https://hdl.handle.net/10419/51953

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 5844

Teams or Tournaments? A Field Experiment on Cooperation and Competition among University Students

Maria Bigoni
Margherita Fort
Mattia Nardotto
Tommaso Reggiani

July 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Teams or Tournaments? A Field Experiment on Cooperation and Competition among University Students

Maria Bigoni
University of Bologna

Margherita Fort
University of Bologna and CHILD

Mattia Nardotto
University of Bologna and Telecom-Paris Tech

Tommaso Reggiani
University of Bologna and IZA

Discussion Paper No. 5844
July 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 5844
July 2011

ABSTRACT

Teams or Tournaments? A Field Experiment on Cooperation and Competition among University Students*

This paper assesses the effect of two stylized and antithetic non-monetary incentive schemes on students' effort. We collect data from a field experiment where incentives are exogenously imposed, performance is monitored and individual characteristics are observed. Students are randomly assigned to a tournament scheme that fosters competition between coupled students, a cooperative scheme that promotes information sharing and collaboration between students and a control treatment in which students can neither compete, nor cooperate. In line with theoretical predictions, we find that competition induces higher effort with respect to cooperation and cooperation does not increase effort with respect to the baseline. However, this is true only for men, while women do not seem to react to non-monetary incentives.

JEL Classification: A22, C93, I20

Keywords: education, field experiments, incentives, competition, cooperation

Corresponding author:

Tommaso Reggiani
Department of Economics
University of Bologna
Piazza Scaravilli, 2
40126 Bologna
Italy
E-mail: tommaso.reggiani@unibo.it

* We thank G. Calzolari, M. Casari, M. Cervellati, A. Ichino, P. Vanin, G. Brunello and seminar participants at the University of Bologna at the Department of Economics and at the Brucchi Luchino 2010 Workshop for their useful comments.

1 Introduction

In the last years a large debate has focused on the possible ways to improve schooling achievement at every level of education. The relevance of this goal is not disputable, since education contributes to the accumulation of human capital, the development of societies and it is considered as one of the main channels for the reduction of inequality. Recent studies have addressed this goal measuring the impact of monetary incentives both on input (e.g. subsidizing the purchase of learning supports) and on output (e.g. giving money based on grades, or conditional on passing the exam). In this paper we take a different approach and we study the effect of different grading rules on schooling achievement. In particular, we do not use monetary rewards to incentivize students' effort whilst we assign them to different incentive schemes: a tournament, a piece rate and a scheme that promotes cooperation. The analysis is performed on a sample of students enrolled in a undergraduate course in econometrics at the University of Bologna (Italy).

The design of the field experiment is based on a theoretical model that contemplates three different incentive schemes. As a benchmark we consider the effect on effort of a piece rate reward. Then we analyze two alternatives: a tournament that fosters competition among matched students and a cooperative scheme in which they can share information and collaborate. The model suggests a weak ordering between the three: in a competitive environment individual performance should be weakly higher than in the benchmark and effort under the benchmark should be weakly higher than in the cooperative scheme. We also show that the detrimental effect of cooperative incentives on effort does not depend on the specific shape of the distribution of types in the population, while the magnitude – but not the sign

– of the effects of a competitive incentive scheme depends on the shape of this distribution. To test these theoretical predictions, we randomly assign students to the treatments and we adopt a between-subjects design, i.e. each subject is only exposed to a single incentive scheme.

Data confirm the theoretical predictions in the full sample. Moreover, we show that an important difference emerges between genders: promoting competition appears to have a strong positive effect on the exerted effort only for males. In contrast, promoting cooperation reduces effort with respect to the case where students can neither compete nor cooperate, but this effect is not statistically significant for both genders. These findings are in line with the literature on how competition affects behaviour depending on gender (see for example Gneezy, Leonard, and List 2009) and provide an interesting comparison with respect to the result of Angrist and Lavy (2009) who find that monetary incentives improve performance especially on girls. We depart from this branch of the literature, complementing the results obtained through monetary incentive, by focusing on non-monetary ones since they represent a relatively cheap way to increase student's effort[1].

The paper proceeds as follows. After a brief review of the related literature (Section 1) we describe and discuss in detail our experimental design (Section 2). In Section 3, we present a simple model, and derive the theoretical predictions which will serve as a reference for the analysis of the experimental data, presented in Section 4. Section 5 concludes, and presents possible extensions of this research.

[1]Studies on monetary incentives proved to be successful in improving students' performance but the cost of inducing higher effort is not negligible. In a study conducted in the New York City school system $600 have been awarded for each passing grade, the Baltimore City Public School District has paid up to $110 to improve scores on state graduation exams and similar programs in the US award up to $500 for each exam passed.

2 Related Literature

We provide a brief review of the literature to highlight our contribution with respect to previous theoretical and empirical work.

On the general issue of how to foster students' effort and school achievement through explicit incentive schemes, several papers explore the role of pecuniary-based incentives. Among those, Blimpo (2010) represents the closest study to our experiment. Analysing data from a field experiment in Benin with a pool of 100 secondary schools, he studies whether individual or different kind of team incentives can lead to a higher students' school performance. He considers three treatments. In the first treatment, each student obtained an individual monetary reward if and only if his or her performance exceeded a minimal threshold at the final exam. In the second treatment, participants were randomly assigned to teams of four students and each team-member received a monetary reward depending on the average team performance, if and only if all the team-members achieved a target performance level. Finally, in the third treatment, participants were randomly assigned to teams of four students but in this case only the components of the three top-performer teams were awarded with a monetary prize. Blimpo (2010) finds that the individual based incentive scheme with cut-off target is most effective for students at an intermediate performance level: at the lower tail of the skills distribution, students reduce effort, probably because they perceive the target out of reach; at the higher tail of the distribution, students know that they are able to get the prize without any extra effort, thus the average impact of such incentives is smaller. When teams are evaluated according to the average performance of the group conditionally on the achievement of a minimal performance target (2nd) treatment),

students across all levels of ability are positively affected: the effort exerted by the different team-mates is pushed toward the target. The tournament scheme (3rd treatment) yields the most beneficial effects: it induces all the teams to work harder as students exposed to this treatment do not have any prior information about the quality and the skills of their competitors in the other teams.

Recent papers consider tournaments at school with financial rewards. Kremer, Miguel, and Thornton (2009) focus their study on the evaluation of a merit scholarship programme dedicated only to female students in an elementary school in Kenya. They observe a substantial increase in the exams scores: in particular girls with low pre-test scores, who were unlikely to win a scholarship (and actually did not get it), reported positive and significant gains in terms of higher school performance. De Paola and Scoppa (2010) studied the effectiveness of monetary incentive schemes in enhancing students' performance using a randomized experiment involving undergraduates in an Italian University. Students participating in the experiment were assigned to three different groups: a high reward group, a low reward group and a control group. Rewards were assigned according to a ranking rule to the top performing students in each treated group. The authors report that financial rewards contributed to increase the students' performance: a very strong reaction emerged among high ability students who were likely to win the contest, while no significant effect was observed for low ability students that have fewer chances to win the tournament competition. Along the same lines, Leuven, Oosterbeek, and van der Klaauw B. (forth.) present results of a randomized field experiment in which freshman students at the Amsterdam University had the opportunity to earn financial rewards for passing all first year requirements. Their findings provide evidence that high ability students perform significantly better when assigned to re-

warded groups. On the contrary low ability students' outcome decreases if assigned to rewarded groups. The small aggregate average effect that they observe is therefore the sum of a positive effect for high ability students and a negative off-setting effect for low ability students. These previous results highlight the importance of controlling for students' ability and individual characteristics when assessing the impact of incentive schemes on their school performance.

A recent study has Fryer (2010) addressed this goal measuring the impact of monetary incentives both on input (e.g. subsidizing the purchase of learning supports) and on output (e.g. giving money based on grades, or conditional on passing the exam). Results show that incentives can raise achievement among even the poorest minority students in the lowest performing schools if the incentives are provided on "'inputs"'. Incentives focused on "'output"' result to be much less effective.

Among the authors who studied the effects of financial incentives on "'output"', some focused specifically on gender differences. Angrist and Lavy (2009) evaluate the effectiveness of financial rewards on the achievement of Israeli students using a randomized experiment providing monetary awards to students who obtain the university admission. The authors show how the program led to significant effects for girls but not for boys. Differences in gender-scheme interaction emerge also from the field experiment by Angrist, Lang, and Oreopoulos (2009). In this study, researchers randomly assigned a sample of students enrolled in a Canadian university to one of three different treatments: the first group was provided with a set of support services (e.g. tutoring); the second group was offered financial rewards for good academic scores; the third one was offered a combination of support services and monetary incentives according to the academic performance. The results of the

experiment show that while males did not react to any of the treatments, females improved significantly their academic performance when monetary incentives were provided.

While females appear to react more than males to monetary incentives awarded for achieving an exogenously given target, incentive schemes based on competition may yield opposite effects. Gneezy, Niederle, and Rustichini (2003) found that males are more prone to engage in competition than females and in general males' performance increases more than the females' one when subjects are exposed to a competitive setting. Similarly, Niederle and Vesterlund (2007) find that, when given the opportunity to choose between a piece-rate payment scheme or a tournament, men select the tournament twice more frequently than women, suggesting that women tend to avoid competition when they have the chance to do so. Azmat and Iriberri (2010) find that, even when the incentive scheme is based solely on the subject's performance, providing information about the *relative* performance promotes higher levels of effort among men, but not among women. We explore the role of gender, and we find that males tend to respond to incentives as predicted by the theory, while females do not.

From a theoretical standpoint, Bratti, Checchi, and Filippin (2008) proposed a model of student cooperation/competition in learning activities, showing that free riding opportunities lead to an insufficient degree of cooperation between schoolmates, which in turn decreases the overall achievement of the group. According to their analysis, a cooperative learning approach may successfully emerge when the class is homogeneous in terms of students' ability. In our study we consider an experimental design and a theoretical model where the incentive scheme is exogenous but similarly to Bratti, Checchi, and Filippin (2008) we focus on student

cooperation/competition in learning activities. Our theoretical model suggests that in a competitive environment individual performance should be higher than in the cooperative environment.

3 The Experimental Design

The experiment involved all the undergraduate students enrolled in the Introductory Econometrics course of the major in Management Studies at the University of Bologna, in year 2010.[2] The course lasted 10 weeks (a three-hour-lecture per week). Students participating to the experiment had to undertake 5 tests whose marks were translated into bonus points for the final exam. The bonus points for the final exam were equal to the average mark the student obtained in the five tests.[3]

Tests have been scheduled every two weeks and each test consisted of five multiple-choice questions to be answered in 50 minutes. Each test concerns all topics taught in the course until the last lecture before the test.

Tests were computerized[4], and were held in the computer laboratory of the School of Economics of the University of Bologna. Desks were arranged so to minimize the possibility for students to talk during the exams (see Figure A.1 in Appendix).

The mark in each test consisted in an individual component, based on the number of correct answers in the test, and a number of extra points related to the treatment and possibly to the score achieved by the partner.

Our study included two treatment conditions – characterized by a competitive

[2]The University of Bologna is considered the oldest University in Europe and counts on average nearly 8000 enrolled students each academic year.

[3]Marks in the final exam range from 0 to 30. The exam is passed with a mark equal or above 18. The bonus points ranged from 0 to nearly 4.

[4]The experiment was programmed and conducted with the software z-Tree (Fischbacher (2007a)).

and by a cooperative incentive scheme, respectively – and a control treatment. In all treatments including the control, part of the incentive depended solely on individual effort. Treatments differed in how tests two, three and four were performed, while the first and the last test were identical across treatments. The first and the last tests were taken individually by each student. In contrast, in the second, third and fourth tests students in the two treatment conditions were randomly matched in couples at the beginning of each test, and had the opportunity of exchanging messages with their partner via a controlled chat program, running on their computer. In both treatment conditions, the total score in tests 2, 3, and 4 of the test depended not only on the student's individual effort (i.e. the net score), but also on the partner's performance. Table 3.1 summarizes the treatments, which are described in detail below.

Students were assigned to treatments between the first and the second test. Before starting tests 2, 3, and 4, students assigned to the two treatment conditions were asked whether they wanted to use the chat or not to communicate with the paired partner. This decision was taken simultaneously by all students. During the test, a couple of students could use the chat program only if both students declared to be willing to communicate, at the beginning of the test. If the two students chose to communicate, per each of the questions of the test they could send only one "signal" to indicate what the right answer was, and one short text message of up to 180 characters. Interactions were anonymous, as students could not know the identity of their partner. In the control treatment no interaction between students was allowed.[5]

[5]Figure A.4 presents a screen-shot of the graphical interface of the program used for the tests On the left-hand side of the screen students could read the question, and the multiple-choice answers. On the top-right part of the screen they could send messages to their partner, while on the bottom-right part of the screen they could read the messages possibly sent to them by the partner.

In each test, the value p^q of correct answers to each question q ranged between 0.3 and 1.2 points. Across all treatments, the number of points v_i^k a student could get by correctly answering the questions of test k was:

$$v_i^k = s_i^k \cdot \mathrm{I}\left(s_i^k \geq 1.5\right), \quad s_i^k = \sum_{q=1}^{5} p_i^{q,k}, \quad k = 1, \ldots 5$$

In each test, the maximum number of points $\bar{v}$ was equal to 3. This is the individual part of the mark in the test, i.e. the component which is common across all treatments.

In the COMPETITIVE treatment, student i's mark in a test was increased by 2 extra points if her score resulted to be strictly higher than the partner. The k-th test's mark $\hat{v}_i^k$ for student i under this incentive scheme is described in equation (1).

$$\hat{v}_i^k = v_i^k + 2 \cdot \mathrm{I}\left(s_i^k > s_j^k\right), \quad k = 2, 3, 4 \tag{1}$$

This provides an incentive for both matched students to compete.

Conversely in the COOPERATIVE treatment, student i's score in a test was increased by 1 extra point if the partner's score was sufficiently good. The k-th test's mark $\hat{v}_i^k$ for student i under this incentive scheme is presented in equation (2).

$$\hat{v}_i^k = v_i^k + \mathrm{I}\left(s_j^k \geq 1.5\right), \quad k = 2, 3, 4 \tag{2}$$

Finally, students in the CONTROL treatment received 1 extra point in tests 2, 3 and 4.[6]

[6]This is done so that the maximum number of bonus points per team is constant across treatments.

Table 3.1: Summary of the treatments, in tests 2, 3 and 4

treatment	**extra points (rounds 2, 3, 4)**	**messages available**
CONTROL	1	no
COOPERATIVE	$\mathrm{I}(s_j^k \geq 1.5)$	yes
COMPETITIVE	$2 \cdot \mathrm{I}(s_i^k > s_j^k)$	yes

Time-line of the experiment. The experiment started in February 2010, and ended in July of the same year. In the first lecture of the course, on February 25th, the full set of instructions was distributed to students and each student had two days to decide whether to take the partial exams or not. At this stage, students were not explicitly informed that they were taking part in an experiment and only at the very end of the course, participating students were asked to sign a consent form authorizing the treatment of data collected during the partial exams.[7]

On March 1st, during a standard class, students were asked to fill in a questionnaire collecting data about some personal characteristics (age, gender, familiarity with computers, e-mail and chat programs, mother and father education). Questionnaire answers are used in the econometric analysis to control for individual-specific characteristics.[8]

On March 22nd students took the first test. Notice that at this stage students had not yet been assigned to treatments, so the grade in this first test can be used as a measure of their effort level before being exposed to the treatment. Students received information about what treatment they had been assigned to only three days later, on March 25th.[9] In the same day, students were informed about their

[7]The experiment was authorized by the ethics committee of the the University of Bologna *(Comitato Bioetico per la Valutazione di Protocolli di Sperimentazione)*.

[8]An overview of the answers to the questionnaire is provided in Section 5, and a translation of the questions is reported in Table A.2 in Appendix.

[9]Students taking part in our experiment were then randomly assigned to two groups of about 65 people each, because the computer lab can host only up to 80 students at a time. All students assigned to the competitive treatment and half of those assigned to the control treatment were in the

own result in the first test, and about the distribution of the first test score among participants. In this way we tried to convey common knowledge of the distribution of competences and ability in the population. Section 4 will show how this is relevant from the theoretical point of view.

The remaining four tests were taken approximately every two weeks, in April and May 2010 with the exception of the fifth which was administered one week after the fourth.[10] Student could benefit of the bonus points gained in the tests only if they took the final exam in June or July 2010. On March 22, before the experiment started, students were informed that the bonus points would expire after the summer.

4 The Model

This section describes the main features of the model we use to derive theoretical predictions and inform the experimental design. After briefly characterizing the general features of the model, we illustrate its implications in terms of expected effort under the different incentive schemes. We first describe what happens without competitive or cooperative incentives (BASELINE treatment). We then characterize the optimal effort under incentives to cooperation and to competition and finally we highlight the testable predictions of the model.

General features We assume that students' abilities are in the interval $\theta \in [0, 1]$ and are distributed according to a non-degenerate distribution function $F(\cdot)$. Stu-

first group, while all students in the cooperative treatment and the remaining students of the control treatment were in the second group.

[10]This is made on purpose since the last test is taken by students individually and covers the last contents of the program as well as some of the previous ones. Hence, it will reflect the effort exerted in the previous stages.

dents choose a level of effort $e_i \in [0, 1]$, which determines their score in the tests and the grade in the final exam. The dis-utility from effort is $c(e_i)$. We further assume that $c(\cdot)$ is independent on subjects' ability θ_i, and that $c'(\cdot) > 0$ and $c''(\cdot) > 0$.

The expected score in test k is a function of ability and effort and is given by the following expression:

$$\mathrm{s}_i^k = e_i \cdot \theta_i \cdot \bar{v} \tag{3}$$

The utility of each student is positively affected by the score and negatively affected by the effort. We assume that students choose their level of effort two times: the first time they choose $e_{i,0}$ when the course starts, before the first test and before the assignment to the treatments; later, after having been assigned to treatments they choose the level of effort e_i that determines their scores in tests 2 to 5 and in the final exam. At this point, their expected utility is given by 3 components: the bonus points obtained in the four remaining tests to be taken – which in the two treatment conditions is the outcome of the interaction with the matched agent – the individual mark in the final exam[11] and the cost of effort. Under the assumption of risk neutrality, the expected utility at the time in which e_i is chosen is:

$$E[U_i] = \frac{1}{5} \sum_{k=2}^{5} \int_0^1 \hat{v}_i^k(\theta_i, e_i, \theta_j, e_j) \cdot f(\theta_j) \mathrm{d}\theta_j + \bar{V} \cdot e_i \cdot \theta_i - c(e_i) \tag{4}$$

where $\bar{V}$ is the maximum mark in the final exam.

[11] Remember that the bonus adds points on top of this mark.

Baseline treatment A student assigned to the baseline treatment does not interact with any other student. As a consequence, considering the four tests and the final exam, the expected utility (4) simplifies in:

$$U_i^{BL} = \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5} \cdot (4 \cdot e_i \cdot \theta_i \cdot \bar{v}) + \frac{3}{5} - c(e_i) \tag{5}$$

from this utility function we can derive the optimal effort exerted:

$$\frac{\partial U_i^{BL}}{\partial e_i} = (\bar{V} + \frac{4}{5} \cdot \bar{v}) \cdot \theta_i - c'(e_i) \tag{6}$$

Normalizing the quantity $V + \frac{4}{5} \cdot \bar{v} = 1$, we get the baseline effort:

$$c'(e_i^{BL}) = \theta_i$$

that implies

$$\frac{\partial e_i^{BL}(\theta_i)}{\partial \theta_i} > 0$$

i.e., we expect more able individuals to exert more effort in the baseline treatment with respect to less able individuals and no variation over the optimal choices of effort.

Competitive treatment To model student's behavior under the two treatments and to derive predictions, we look for the equilibrium in the Bayesian-Nash games where students have private information about their own type and a common knowledge on the distribution of abilities in the population.

Under the competitive scheme, students get bonus points if their score is higher

than the partner's. Equation (7) describes the expected utility in this case.

$$
\begin{aligned}
U_i^{comp} &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{3}{5} \cdot 2 \cdot Pr(e_i \cdot \theta_i > e_j \cdot \theta_j) - c(e_i) = \\
&= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{6}{5} \cdot \int_0^{\theta_i \cdot \frac{e_i}{e_j}} f(\theta_j) d\theta_j - c(e_i) = \\
&= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{6}{5} \cdot F\left(\theta_i \cdot \frac{e_i}{e_j}\right) - c(e_i)
\end{aligned}
\tag{7}
$$

where $6/5 \cdot F(\theta_i \cdot e_i/e_j)$ is the expected number of additional points obtained in the second, third, and fourth test in case the student outperforms his partner. Hence, the expected utility can be expressed as:

$$
e_i(\theta_i) \in \underset{e_i}{\text{argmax}} \left\{ E[U_i] = \bar{V} e_i \theta_i + \frac{1}{5}\left[4 e_i \theta_i \bar{v}\right] + \frac{6}{5} \int_{\theta_j | \theta_j e_j < \theta_i e_i} f(\theta_j) d\theta_j - c(e_i) \right\}
\tag{8}
$$

Under regularity assumption on the distribution of types in the population, it can be shown that the first order conditions are[12]:

$$
\theta_i - c'(e_i) + \frac{6}{5} f(\Phi_j(e_i)) \Phi'(e_i) = 0 \tag{9}
$$

where Φ_k is the mapping from the effort to the type (individual ability). Now, since $\Phi' = 1/e'$, we have the following solution for the optimal effort in the competitive treatment:

$$
c'(e_i) = \theta_i + \frac{6}{5} f(\theta_i) \cdot \frac{1}{e_i'} \tag{10}
$$

[12]In order to have a pure strategy Nash equilibria, the distribution function of types must be non-degenerate and the mapping from type to effort must be continuous and increasing. The requirement on the distribution of types is a plausible requirement, given the heterogeneity in the population while the two on the mapping between type and effort can be proven to be true in our case. In the non-heterogeneous case, that is when the distribution of types is degenerate, it can be easily shown that no pure-strategy equilibrium exists.

From this equation we see that the optimal effort exerted under this scheme is equal or higher than the optimal level of effort e_i^{BL} in the control treatment. The magnitude of the effect depends on the subject's ability θ_i and on the shape of the distribution $F(\cdot)$.

Cooperative treatment Under this scheme, each student has a clear incentive to share her information (in tests 2, 3 and 4) and the mark depends also on the partner's effort.

In this case the expected utility becomes:

$$\begin{aligned} U_i^{coop} &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5} \cdot [e_i \cdot \theta_i \cdot \bar{v}] + \\ &+ \frac{1}{5} \int_0^1 3 \cdot [\bar{v} \cdot (e_i \cdot \theta_i + e_j \cdot \theta_j - e_i \cdot \theta_i \cdot e_j \cdot \theta_j) + \\ &+ \mathrm{I}(e_i \cdot \theta_i + e_j \cdot \theta_j - e_i \cdot \theta_i \cdot e_j \cdot \theta_j > 0.5)] \cdot f(\theta_j) \mathrm{d}\theta_j - c(e_i) \end{aligned} \tag{11}$$

The second term in equation (11) represents the points obtained form the fifth test, where no interactions among student was allowed, while the third term represents the bonus obtained in tests 2, 3 and 4.

The assumption that information is shared by the students is crucial and implies that the probability of knowing the answer is given by the common knowledge of the couple. Thus, here the knowledge of the couple is the union of the knowledge of the two members and the optimal effort is given by:

$$c'(e_i^{coop}) = \theta_i - \frac{3}{5} \cdot \bar{v} \cdot \theta_i \int_0^1 \theta_j \cdot e_j \cdot f(\theta_j) \mathrm{d}\theta_j \tag{12}$$

The second term in the right-hand side of equation (12) is always non-positive, and its absolute value increases with θ_i. This shows that, since information is shared,

each team member has an incentive to exploit the effort of the other lowering his own contribution. As a consequence, under the cooperative treatment, team members have an incentive to shrink their effort, and this detrimental effect of cooperation on effort is stronger for students with higher ability (θ_i).

Testable predictions To sum up, our theoretical model predicts that, given the ability θ_i, the effort exerted by student i in the three treatments is such that:

$$e_i^{coop} \leq e_i^{BL} \leq e_i^{comp}$$

i.e., we expect that on average students randomized into the COOPERATIVE treatment exert lower or equal effort than students randomized into the CONTROL treatment whereas students randomized into the COMPETITIVE treatment should exert more effort.[13] Conversely, at test 1, all students have the same individual incentives to increase effort and optimal effort depends only on their ability level, i.e. $e_{i,0}^{coop} = e_{i,0}^{BL} = e_{i,0}^{comp} = e_{i,0}$. Moreover, the model predicts that the detrimental effect of the cooperative scheme is stronger for high ability individuals while the same type of individuals should exert more effort with respect to the less able individuals in the baseline treatment. Note that our main testable predictions involve the differential changes in effort across treatments and ability levels. Our design allows to measure these changes, as discussed in more detail in section 5.1.

We also expect that students assigned to the cooperative treatment will use the chat more frequently and will use it to exchange information. Conversely, students assigned to the competitive treatment should use the chat less frequently and could potentially use it for acts of sabotage, i.e. to suggest the wrong answers. We col-

[13]The ordering holds if the distribution of abilities is the same in the three treatments.

lected data to check these aspects. Results of our inquiry are discussed in section 5.3.

5 Results

In this section we first discuss our choice of outcome measure, then present the data and discuss the results on the effect of the incentives on information sharing and on effort.

5.1 Measuring Effort

Our theoretical model predicts that for a given level of ability, there is a weak ordering in the effort exerted by each student i, namely $e_i^{coop} \leq e_i^{BL} \leq e_i^{comp}$. We thus expect that on average students randomized into the COOPERATIVE treatment exert lower or equal effort than students randomized into the CONTROL treatment whereas students randomized into the COMPETITIVE treatment should exert more effort.[14]

Equation (3) in our simple model describes the relationship between expected student's score at each test and effort, namely $s_i = \theta_i e_i$, where s_i is the net score of individual i, θ_i is a measure of individual ability and e_i is the effort exerted.

Taking logs and allowing for noise in the way in which effort generates the score, we get

$$y_i = \zeta_i + \epsilon_i \tag{13}$$

where $y_i \equiv log(s_i)$ is the log of the net score of individual i, $\zeta_i \equiv log(e_i)$ is the log of the effort exerted, while $\epsilon_i = log(\theta_i) + \varepsilon_i$ and $E[\epsilon_i] = log(\theta_i)$, i.e. we assume

[14]The ordering holds if the distribution of abilities is the same in the three treatments and this is guaranteed by randomization.

that only the idiosyncratic component ε averages to 0 for any i, while the error ϵ_i has a possibly non-zero mean equal to an individual specific constant.

Our experimental design provides an interesting way to measure effort under weak assumptions. Recall that we observe students' performance in similar tests both before the assignment to the treatments (test 1) and after the exposure to the treatments (test 5). Both these tests are taken individually under all treatments and cover similar topics[15]. However, by construction, the score in the first test and the effort exerted to pass it cannot be affected by the treatments since both performance and effort are pre-determined with respect to the assignment to the different incentive schemes. Conversely, the score in the last test should reflect changes in effort induced by the treatment. Indeed, moving from equation (13) and contrasting the performance in test 5 and 1, we have $y_i - y_0 = \zeta_i - \zeta_{i0} + \varepsilon_i - \varepsilon_{i0}$. It follows that $E[y_i - y_{i0}] = E[\zeta_i - \zeta_{i0}]$, i.e. by looking at the change in the logarithm of score between the first and last test, we measure the change of the logarithm of effort net of the direct effect of any fixed individual specific factor.

Recall that all our treatment conditions have a **common** individual incentive to increase effort but differ in the incentives to compete or cooperate and only in the baseline students can neither compete, nor cooperate. Following the theoretical predictions of our simple model, we expect an increase in effort in all treatments with respect to a set up where no individual incentives are granted. Our experiment is not designed to estimate this common effect -none of our groups has no individual incentives- but to capture the differential changes induced by the different treatments. The testable prediction of our model involves the differential increase

[15]The last test covers a larger set of arguments which includes also those covered by the first and is more closely spaced over time with respect to the other tests.

in effort under the cooperative and competitive scheme with respect to the baseline. This weak ordering holds also if we consider $log(e)$, since the logarithm is a monotonic transformation.

To test the theoretical predictions, we first contrast the distribution of effort under the three schemes and check for heterogeneity in the treatment effect over the effort distribution. We then assess the effect on the average change in $log(e)$ and run the following regression

$$E[\zeta_i - \zeta_{i0}] = \beta_0 + \beta_1 Coop + \beta_2 Comp \tag{14}$$

where β_0 represents the average change in $log(e)$ under the baseline, β_1 is the average differential change in $log(e)$ under the cooperative scheme with respect to the baseline, and β_2 is the average differential change in $log(e)$ under the competitive scheme with respect to the baseline. The theory predicts $\beta_1 \leq 0$ and $\beta_2 \geq 0$. There is an additional prediction that $\beta_0 = 0$, i.e. no change in effort under the baseline. However, our model does not allow for learning which may occur in practice. Namely, after the first test the score of the students in the baseline improves because they are becoming more familiar with the types of tests and the way the tests are performed in the laboratory. Allowing for learning will not affect our theoretical predictions provided that learning is constant across treatments. If learning occurs in practice, $\beta_0 > 0$.

5.2 Data and Descriptive Statistics

Among the 145 students attending the course, 131 applied for participation into the experiment. Our elaborations are based only on the records of the *stayers*, i.e. 114

students who participated to all 5 tests.

We exclude from the elaborations the records of 17 students who missed at least one test: 10 students assigned to the control treatment (BASELINE in what follows), 2 students assigned to the COOPERATIVE treatment and 5 students assigned to the COMPETITIVE treatment (see table A.1 in Appendix). We shall highlight that 6 of these students were late at the 3rd test and were thus excluded from that test. The experimental program is run in z-Tree Fischbacher (2007b): when the test (the experimental session) starts, additional subjects can participate only shutting down and restarting the entire session. Students were informed that not being on time for the test would result in being excluded from the test session. Out of these 17 students, 8 dropped out after the first test: all these students were assigned to the baseline treatment after test 1. When we compare stayers and dropouts in the full sample, we cannot reject the null that drop-outs had a worse performance in the first test.[16]

Once we limit the analysis to the students who participated at all tests, the samples are relatively balanced across treatments with respect to observed and pre-determined characteristics: we do not detect differences in the distribution of the score the first test (score 1) and the average score at previous exams (GPA) between any two treatments (BASELINE, COOPERATIVE, COMPETITIVE) at any conventional level of confidence (see Table 5.1). Figures A.2 and A.3 in Appendix report the empirical probability distribution of the pre-treatment variables (the score

[16]There are no significant differences between the subpopulation of excluded students and the stayers in observable and pre-determined characteristics among the students who where assigned to the COOPERATIVE treatment. We do not reject the null of equal means at 1% level -but we reject at 5%- in the subpopulations for the other treatments: students who participated to all tests in the BASELINE and in the COMPETITIVE treatment tend to be those who achieved a higher score in the first test (0.7 points higher than the one for those who dropped out in the BASELINE group and 0.85 points higher than the one for those who dropped out in the COMPETITIVE treatment).

in the first test, and the average mark in previous exams). Table 5.1 also reports the mean value of several other individual characteristics, obtained from subjects' answers to the questionnaire and p-values of tests aimed at detecting differences in these characteristics across treatments.[17] In general, the overall sample is well balanced across treatments. There are some exceptions: the frequency of use of e-mail is significantly higher in the BASELINE treatment than in the COMPETITIVE and in the COOPERATIVE treatments. Significant differences emerge also in terms of the education level achieved by the students' fathers (but not mothers).

To detect the role of interactions effect between the treatments and the students' ability, we consider several different proxies for student's ability and include interaction terms in a simple regression. Our favorite proxy to control for student's ability is the average mark at previous exams: students who participated in the experiment are third year students taking exams is in the last quarter of the third year; therefore, their academic history can be a reliable proxy of their academic skills. In line with the most recent empirical evidence from Italy (AlmaLaurea, 2009), also in our sample females tend to perform significantly better than males in terms of GPA (Females $= 25.2$, Males $= 24.3$, Rank-Sum Test $= P\ 0.028$). We say an individual is a high ability individual if his/her score on the classification variable is above the median for that variable in the sample.

5.3 Communication and treatments.

Students under both treatments' schemes had two ways to communicate: they could send text messages or hints[18]. Messages and hints were limited in two ways. On

[17] We contrasted averages across treatments by means of linear and non linear regressions.

[18] The hint consisted in a simple message informing the receiver that the sender believes a certain answer to be the right one. The sender can suggest a different answer with respect to the one actually selected in the test.

characteristic	pooled	mean baseline (BL)	cooperative (COOP)	competitive (COMP)	p-values BL vs. COOP	BL vs. COMP	COOP vs. COMP
GPA	24.8	24.8	24.9	24.8	0.808	0.883	0.928
score 1	1.8	1.8	1.9	1.7	0.520	0.564	0.220
age	21.7	21.6	21.9	21.5	0.280	0.736	0.161
gender (Male)	47.4%	40.5%	51.2%	50.0%	0.346	0.418	0.915
freq. mail^{+}	46.7%	62.9%	43.2%	33.3%	0.098	0.017	0.396
freq. chat^{+}	54.3%	54.3%	51.4%	57.8%	0.803	0.785	0.602
freq. pc^{+}	43.8%	45.7%	40.5%	45.5%	0.658	0.983	0.678
father edu.$^{++}$	31.4%	42.9%	13.5%	39.4%	0.008	0.772	0.017
mother edu.$^{+++}$	29.5%	37.1%	24.3%	27.3%	0.241	0.386	0.778
risk aversion	6.0	6.0	6.1	5.8	0.964	0.497	0.515
risk averse	50.9%	48.6%	53.7%	50%	0.659	0.908	0.749
trust 1	4.9	4.7	4.9	5.0	0.567	0.513	0.744
truster (1)	39.5%	27.0%	43.9%	47.2%	0.124	0.077	0.770
trust 2	3.8	3.7	3.8	4.0	0.863	0.533	0.664
truster (2)	27.2%	21.6%	29.3%	30.6%	0.441	0.386	0.902

Table 5.1: Mean value of individual characteristics, by treatment. An individual is risk averse or truster if his answer on the scale is higher or equal to 6. $^{+}$ Binary indicator for whether chat, pc or e-mail are used frequently, i.e. more than once a day. $^{++}$ Binary indicator for whether the father as high education (i.e. if his qualification is equal to college or higher). $^{+++}$ Binary indicator for whether the father as high education (i.e. if his qualification is equal to college or higher). The signifcance of differences across treatments is estimated by means of simple linear and non linear regressions (logit) for binary indicators. P-values are reported.

Table 5.2: Use of the chat

Treatment	Acceptance of the chat	Av. num. of messages	Av. message length
Cooperative	98% of subjects	3 (out of 5)	28 words
Competitive	70% of subjects	0.5 (out of 5)	11 words

the one hand students could not send any information useful to identify themselves (under the threat of exclusion from the test); on the other hand, for each of the 5 questions asked in a test, a student can send and receive only one message of both types.

Table 5.2 together with Table A.3 in the Appendix report descriptive statistics on the use of chat by subjects. The figures suggest that almost everybody under the COOPERATIVE treatment accepted it[19], and that the average number of exchanged messages is six times higher than in the COMPETITIVE treatment.

The chat tended to be used more frequently than the hint under both schemes.

The content of conversations suggests the chat has been actually used to exchange information. Conversely, the chat was not actively used by students under the COMPETITIVE scheme: they declared to be willing to use the chat but only 0.5 messages were exchanged on average. More importantly, students did not believe in the messages of the partner[20]. Indeed, in some cases the chat has been used to deceive the partner (see Table 5.4, and Figure A.6 in Appendix for an illustrative example).

Table 5.3 reports descriptive statistics on the number of actions taken by students under each treatment. Sending a text message or giving a hint are actions. Under the COOPERATIVE scheme the average number of actions tend to increase

[19] At the beginning of the exam the student must input the registration number and then choose if she wants to use the chat or not.

[20] We do not provide descriptive statistics on the extent of sabotage because these statistics would not be comparable across treatments. Indeed, given the low number of individuals that used the chat under the competitive treatment, we will not get reliable statistics for that group.

from the first test in couples (test 2) to the last (test 4), changing from nearly 5 to above 6, and the correlation between the number of actions taken in different tests is positive, between 0.34 and 0.53, and decreasing with the lag between tests. Some students under the COOPERATIVE scheme used all the available actions (5 text messages and 5 hints) and the median number of action is between 6/7: students tended to use at least one of the two available actions in each question of each test and they often used both. Generally, the text message was sent before the hint, and the time lag between the text message and the hint ranges between 1 and 5 minutes in most questions and tests (see Table A.3 in the Appendix). Conversely, under the COMPETITIVE scheme the median number of actions taken is always 0 and the average number of actions remains relatively stable slightly above 1: students tend to use both the chat and the hint for the same question and only once per test. They also tend to send the text message and the hint almost simultaneously or to send the hint before the text message (see Table A.3 in the Appendix). The correlation between the number of actions taken in subsequent tests is weaker (between 0.17 and 0.36) and tends to increase with the lag between tests. The correlation between the exerted effort and the number of actions is negligible under both schemes.

We consider data on the couples in each test and contrast answers of the members: Table 5.4 shows that members of the couples under the COOPERATIVE scheme tend to give the same answer much more frequently than their class mates under the COMPETITIVE scheme. The difference is stable across tests and slightly higher than 25%.

We interpret the observed pattern of information exchange across treatments as a positive response to the incentives: students understood the different mechanisms underlying the two different schemes and behaved accordingly as far as exchange

Table 5.3: Number of actions (i.e. use of chat and use of hints) by round and treatment.

	Cooperative				
	mean	sd	median	min	max
Test 2	5.12	3.36	6	0	10
Test 3	5.80	2.92	7	0	10
Test 4	6.37	2.91	6	0	10
	Competitive				
	mean	sd	median	min	max
Test 2	1.47	2.48	0	0	8
Test 3	1	2.51	0	0	10
Test 4	1.67	2.24	0	0	8

Table 5.4: Proportion of cases in which the members of the couple give the same answer.

	Test 2	Test 3	Test 4
Cooperative	56.38%	77.26%	84.78%
Competitive	30.5%	52%	56.84%
Difference	25.88	25.26	27.94

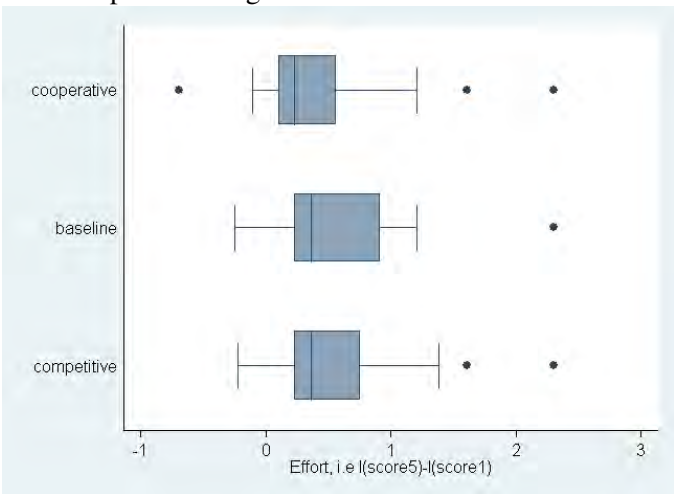
of information is concerned.

5.4 Treatment effects

Figure 5.1 depicts the empirical distribution of effort (i.e. log(net score 5) - log(net score 1)) across treatments. The vertical blue line represents the median of the distribution, the left hinge of the box indicates the 25th percentile, and the right hinge of the box indicates the 75th percentile. Visual inspection suggests that under the COOPERATIVE treatment, subjects perform more poorly respect to the BASELINE treatment, while no sizable differences emerge between the COMPETITIVE and the BASELINE treatments.

Wilcoxon tests do not reject the hypothesis that the distribution of effort is the same across treatments. These tests are not appropriate if we want to establish an

Figure 5.1: Box-plot showing the distribution of effort across treatments.



ordering across all three treatments. Thus, we also perform a Jonckheere-Terpstra test, a non-parametric test designed to detect alternatives of ordered class differences. This test does reject the hypothesis that effort is constant across treatments versus the alternative hypothesis that effort is ordered across treatments according to our main theoretical prediction ($e_i^{coop} \leq e_i^{BL} \leq e_i^{comp}$) at 10%.

P-values of these tests are reported in Table 5.5, together with the mean level of effort in each treatment condition.

Table 5.5: Mean level of effort, by gender and treatment.

	pooled	**males**	**females**
Mean effort			
cooperative	0.500	0.377	0.628
baseline	0.583	0.452	0.677
competitive	0.570	0.680	0.459
Wilcoxon tests (p-values)			
base. vs. coop.	0.313	0.135	0.948
base. vs. comp.	0.745	0.442	0.721
coop. vs. comp.	0.190	0.059*	0.713
Jonckheere-Terpstra tests (p-values)			
	0.088*	0.016**	0.624

Legend: One star, two stars, three stars for significant differences at 10%, 5% and 1% level respectively.

It has been pointed out in Section 2 that according to the experimental literature, a competitive environment may induce different effects on effort for females and for males. Consistently with these works, we find that the picture indeed changes when we split the sample by gender. Figure 5.2 reveals that the treatment effect is substantially different for male and female subjects. The detrimental effect of the COOPERATIVE treatment on effort with respect to the COMPETITIVE treatment only emerges for males, whereas for females no clear treatment effect arises.

One-sided Wilcoxon tests confirms that males' level of effort is significantly lower in the COOPERATIVE treatment than in the COMPETITIVE treatment at 10% level but no significant difference emerges with respect to the baseline. In contrast, the same test does not reject the hypothesis of equal distribution of effort between any two treatments for the female sample. These tests are not appropriate if we want to establish an ordering across all three treatments. Thus we run the the Jonckheere-Terpstra test for the subsamples of males and females: for the male sample, the test rejects at 5% the null hypothesis that effort is not ordered across treatments against the alternative hypothesis that effort is ordered according to what predicted by the theory; no effect is detected for females. P-values of these tests are reported in Table 5.5.

Our theoretical model predicts heterogeneity in the effect of the incentives' schemes on effort with respect to students' ability, at least for the competitive treatment. To control in a parsimonious way for individual ability, and for other individual characteristics, while assessing the effects of the treatments' scheme on average effort, we use linear regression models.

We run the analysis separately for males and females as previous results suggest that they react differently to incentives.

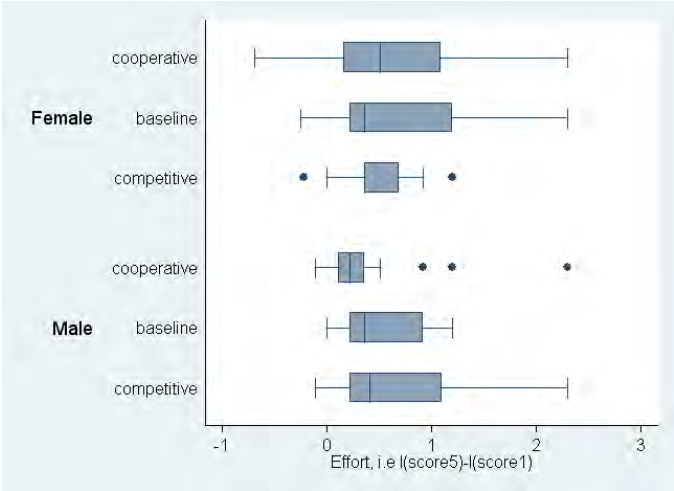


Figure 5.2: Box-plot showing the distribution of effort across treatments, by gender.

Table 5.7 presents the benchmark results of two baseline specifications for males and females: column (1) and (2) do not allow for heterogeneity in the treatment effects with respect to students' ability while in column (3) and (4) we include interactions between treatments and the ability indicator based on the average mark at previous exams. All regressions include controls for father education, risk aversion and trust. The top panel of Table 5.7 reports coefficients estimates while the bottom report p-values of both bilateral and unilateral tests: by specifying the direction in which the null hypothesis of no effect is violated (as predicted by theory), we increase the power of the t-test to detect significant deviations. As reference, we computed the power of a test to detect differences between any two treatments for males and females separately using the descriptive statistics (mean, standard deviations and sample size) of our sample and significance level $\alpha = 0.5$.[21] Table 5.6 shows that we have little power to detect differences between the baseline and the cooperative treatment, while we have more power to detect differences between the baseline and the competitive treatment for both genders, even all values are

[21]The power of similar tests for the pooled sample is lower: the gender heterogeneity makes point estimates of the average $log(e)$ less precise.

Table 5.6: Power of the two-sided and one-sided test of mean comparison across treatments and optimal size n^* for two sided tests with equal group sizes and power 0.8 . Males and Females.

Null Hypothesis	Males			Females		
	Power		n^*	Power		n^*
	2 sided	1 sided		2 sided	1 sided	
Baseline vs Cooperative	0.08	0.12	658	0.06	0.08	3078
Baseline vs Competitive	0.25	0.35	84	0.25	0.36	101
Cooperative vs Competitive	0.07	0.11	879	0.16	0.25	161

quite low. In addition, the table reports the sample size required for a test to detect difference of the size we observe with power 0.8 (assuming constant sample sizes across groups): most of these sizes can be hardly met within a design structured as ours.

Results in Table 5.7 confirm previous results on the differential effects across treatments: there is evidence of a significant increase in effort under the competitive treatment with respect to the baseline for males but not for females.[22] The effect is statistically distinct from zero at 10% and not-negative at 5%. When we control for ability, we find that : (i) the positive incentive for males is higher for the low ability individuals (still significantly non-negative at 10%) and decreases substantially for high ability individuals; (ii) there is a negative and statistically significant (at 10%) detrimental effect of the cooperative treatment for high ability individuals only. However, the difference in effects of incentives between ability groups is not significant in our sample for the competitive case nor for the cooperative case. The magnitude of the effect ranges from 33% to 49% which is a strong increment of the exerted effort. Notice that this is in line with the findings of Blimpo (2010) who use monetary incentives based on the achievement of a specified score target.

For females, no statistically significant effect can be detected. The pattern of

[22] Since we include control variates and 9 students do not answer the questionnaire, the sample size relevant for the regressions is 105 instead of 114.

Table 5.7: Ordinary Least Squares Estimates of the Treatment Effects: Benchmark Specification. Males and Females.

	(1)	(2)	(3)	(4)
	No heterogeneity with ability		Heterogeneity with ability	
Variables	Males	Females	Males	Females
Constant	0.386**	0.652***	0.239	0.740**
	[0.174]	[0.178]	[0.262]	[0.294]
Cooperative	-0.125	-0.156	0.100	-0.135
	[0.178]	[0.219]	[0.263]	[0.382]
Competitive	0.331*	-0.235	0.492*	-0.363
	[0.189]	[0.211]	[0.272]	[0.359]
Coop · High Ability			-0.486	-0.117
			[0.401]	[0.475]
Comp · High ability			-0.266	0.080
			[0.398]	[0.449]
High ability			0.127	0.073
			[0.280]	[0.319]
High parental education	-0.249	-0.155	-0.317*	-0.219
	[0.151]	[0.185]	[0.173]	[0.210]
Frequent use of e-mail			0.139	-0.224
			[0.161]	[0.206]
Risk averse	0.290*	0.221	0.290*	0.280
	[0.147]	[0.173]	[0.150]	[0.198]
Truster (1)	0.105	0.084	0.136	0.114
	[0.155]	[0.194]	[0.164]	[0.213]
Observations	50	55	50	55
R^2	0.237	0.066	0.311	0.157
P-values for the null of no effect against bilateral or unilateral H_1				
(R)$\equiv H_1$: $\beta > 0$; (L) $\equiv H_1$: $\beta < 0$				
Competitive				
1 sided (R)	0.039**	0.867	0.035**	0.844
2 sided	0.089*	0.266	0.071*	0.312
Cooperative				
1 sided (L)	0.240	0.238	0.649	0.362
2 sided	0.480	0.476	0.703	0.724
Competitive for high ability				
1 sided (R)			0.221	0.837
2 sided			0.441	0.326
Cooperative for high ability				
1 sided (L)			0.092*	0.186
2 sided			0.184	0.373

High parental education is a binary indicator that takes the value 1 if the highest qualification of at least one of the parents of the individual is above high school and 0 otherwise. **Risk averse** is a binary indicator that takes the value 1 if the answer of the individual on the risk aversion scale is above 6 and 0 otherwise. **Truster (1)** is a binary indicator that takes the value 1 if the answer of the individual on the trust 1 scale is above 6 and 0 otherwise. Standard errors in brackets. Three stars, two stars and one star for significant effect at 1%, 5% and 10% level respectively.

the effect of competitive incentives on effort for females is similar to the one detected for males but in the opposite direction: the point estimate of the effect is negative and, when we control for ability, point estimates of the effect of competitive incentives for females are negative for both low and high ability individuals but less so for high ability individuals.

We detect a significant increase in effort also in the baseline: we attribute this to the fact that students become more familiar with the instruments used for the test (*learning*). Students' ability does not play any role in determining the increase in effort in the baseline. Few regressors are relevant in determining changes in students effort: risk aversion and parental background attract significant coefficients in some specifications, suggesting that individuals who are risk averse tend on average to increase effort, while males with higher socio-economic background (here proxied by highly educated parents) tend to decrease effort, other things equal.

6 Conclusions

Our study investigates how two alternative incentive schemes affect students' effort, both from a theoretical and from an empirical point of view. To test the theoretical predictions, we run a field experiment in an undergraduate course at the University of Bologna (Italy). We randomly assign students to either a tournament, where coupled students compete to get the reward, a cooperative scheme where information sharing is allowed, or a control treatment in which students can neither compete, nor cooperate. Differently from previous studies, none of our treatments involves pecuniary incentives but consists in extra points for their final grade. By doing so, we provide incentives to students in "the same currency" in which they are usu-

ally rewarded. In this sense, our approach is closer to the literature of personnel economics where workers may receive a salary according to some schedule (based on relative or absolute performance). Hence, we build a bridge between the literature in education economics and personnel economics, showing how dynamics observed in a labor environments arise also in school environments.

The field-experiment data we collected confirm the theoretical predictions: we observe a weak ordering between the effort exerted by students under the different treatments with students in the competitive treatment exerting on average more effort with respect to students in the baseline and in the cooperative treatment.

We break down our results by gender and show that a significant difference emerges: only males react to incentives to compete while we cannot detect significant effect for females. Cooperation seems not to foster effort exertion and no gender effect emerges. In contrast with theoretical predictions we find that students' ability plays little role in determining the effectiveness of the incentives.

Our experimental results suggest that non-pecuniary incentives based on competition have the potential to increase students' effort as pecuniary incentives do (see Blimpo 2010) but at a much lower financial cost. In our case competition proves to work on males which is line with findings in several other contexts (see for example Gneezy and Rustichini 2004 and Niederle and Vesterlund 2010) where it has been shown that males are more prone to compete with respect to females. The estimated increase in effort induced for males in the competitive treatment ranges from 33% to 49%, meaning that, for example, if a student in the baseline spends 3 afternoons in preparing the test (roughly 10 hours), a student under the competitive scheme will spend one more afternoon. Moreover, highlighting the different effect of incentives to compete depending on gender, we complement the

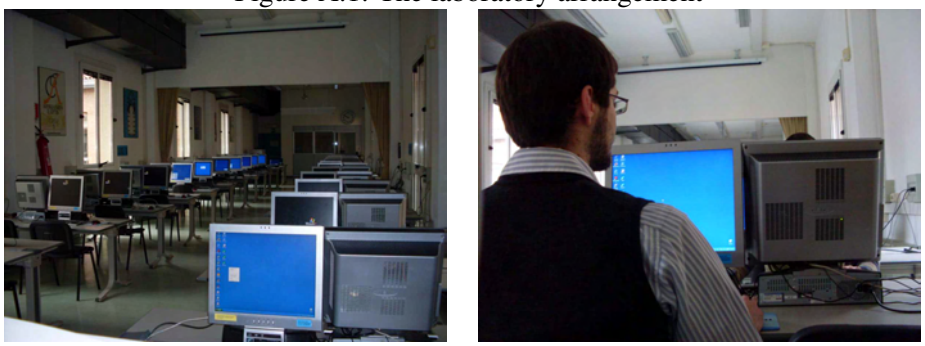
results in Angrist and Lavy (2009) who show that monetary incentives based on absolute performance are more effective for females.

Our study represents a first exploration of the effects of non-monetary incentives on students' performance and effort. It would be interesting to extend the inquiry to different samples, to verify whether our result holds for students with different majors (such as literature of philosophy), who are probably less trained to optimization, and for younger students at high school and middle-high school.

A Appendix

A.1 Laboratory

Figure A.1: The laboratory arrangement



A.2 Additional tables

Table A.1: Descriptive statistics

			Descriptive Statistics- Stayers		
			Predetermined controls		
	assigned	stayers	score 1	exams' avg	score 5
Baseline (control)	47	37	1.80 *(0.81)*	24.76 *(1.8)*	2.91 *(0.24)*
Cooperative	42	41	1.92 *(0.84)*	24.88 *(2.3)*	2.80 *(0.50)*
Competitive	41	36	1.69 *(0.74)*	24.83 *(1.6)*	2.69 *(0.53)*
Full sample	130	114	1.81 *(0.80)*	24.83 *(1.9)*	2.80 *(0.45)*

Score 1: score at the first mock exam. Score 5: score at the last mock exam. Exams' avg: average score at previous exams. Stayers: students who participated to 5 experimental sessions.

In Table A.2, we report the precise definition of questionnaire data used in the analysis.

Table A.2: Description of questionnaire data.

Variable	**Corresponding question**	**Range**	**Coding**
gender	*gender*	0, 1	1 = male
age	*age*	0-100	age in years
freq. mail	*how frequently do you check your e-mail?*	1-5	1="more than once per day" 2="at least once per day"
freq. pc	*how frequently do you use the pc to study/work?*	1-5	3= "at least once per week" 4="less than once per week"
freq. chat	*how frequently do you exchange text messages via chat (msn, facebook, google talk, skype, etc.)?*	1-5	5="Never"
father edu.	*please, indicate the education level achieved by your father*	1-5	1="junior high school" 2="high school"
mother edu.	*please, indicate the education level achieved by your mother*	1-5	3="bachelor" 4="master" 5="Ph.D."
risk aversion	*I would describe myself as a risk-averse person.*	1-10	1="fully agree" 10="fully disagree"
trust 1	*Do you think that most people try to take advantage of you if they got a chance or would they try to be fair?*	1-10	1="people would try to take advantage" 10="people would try to be fair"
trust 2	*Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?*	1-10	1="you can never be too careful" 10="most people can be trusted"

Table A.3: Descriptive statistics -mean, [median] and (standard deviation)- on lag between the use of chat and use of hints, by treatment and round. Questions 1-5

Lag & proportion of user of both chat and hint (seconds). Test 2.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	107.6 [21.7]	322.8 [252.2]	58.6 [6.3]	151.0 [114.1]	54.2 [6.4]
	(449.0)	(378.3)	(565.9)	(513.5)	(495.6)
Users (count)	14	12	11	13	16
Users (%)	35.0%	30.0%	27.5%	32.5%	40.0%
Competitive	80.6 [54.1]	-123.3 [-123.3]	130.4 [23.9]	13.3 [13.3]	25.3 [16.9]
	(67.9)	(131.1)	(1113.5)	(n.a.)	(21.3)
Users (count)	3	2	3	1	3
Users (%)	8.8%	5.9%	8.8%	2.3%	8.8%

Lag & proportion of user of both chat and hint (seconds). Test 3.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	492.5 [374.8]	496.2 [368.1]	-56.0 [-3.2]	167.6 [71.7]	76.4 [5.2]
	(693.1)	(577.4)	(423.0)	(427.3)	(303.5)
Users (count)	14	16	14	17	15
Users (%)	35.0%	40.0%	35.0%	42.5%	37.5%
Competitive	73.8 [7.6]	-164.2 [-164.2]	720.1 [720.1]	-194.0 [165.4]	97.0 [97.0]
	(131.3)	(951.9)	(223.6)	(380.5)	(123.7)
Users (count)	3	2	2	4	2
Users (%)	8.8%	5.9%	5.9%	11.7%	5.9%

Lag & proportion of user of both chat and hint (seconds). Test 4.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	146.9 [55.9]	119.8 [20.8]	-15.8 [-3.3]	169.7 [40.8]	95.0 [4.1]
	(482.1)	(342.8)	(522.4)	(355.0)	(240.2)
Users (count)	14	17	17	17	22
Users (%)	35.0%	42.5%	42.5%	42.5%	55.0%
Competitive	194.3 [12.3]	8.5 [8.5]	180.1 [2.1]	458.8 [72.6]	322.3 [322.2]
	(365.8)	(n.a.)	(314.1)	(721.4)	(449.4)
Users (count)	4	1	3	3	2
Users (%)	11.7%	2.9%	8.8%	8.8%	5.9%

A.3 Additional figures

Figure A.2: Empirical probability distribution of score 1 by treatment

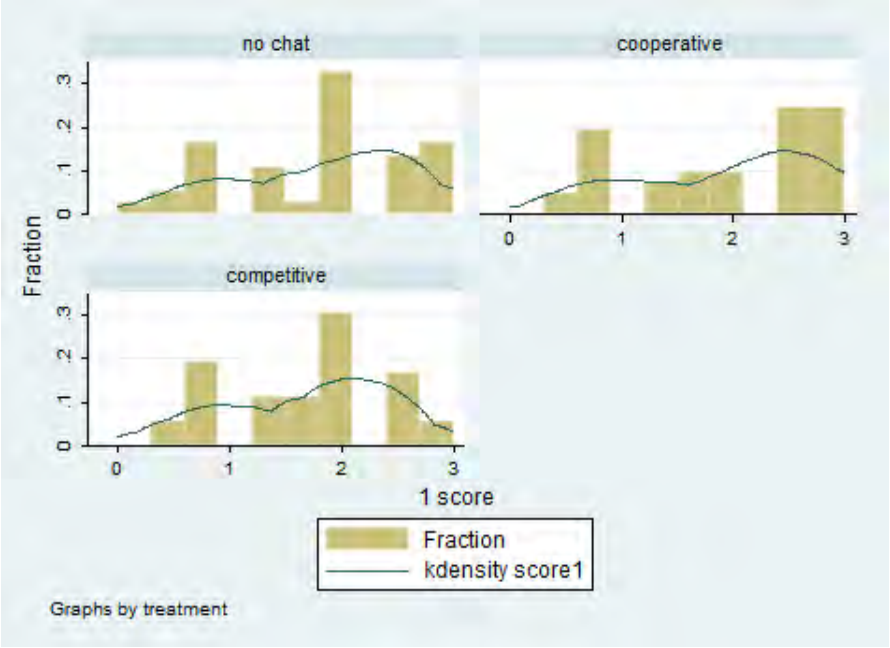


Figure A.3: Empirical probability distribution of average score at previous exams by treatment

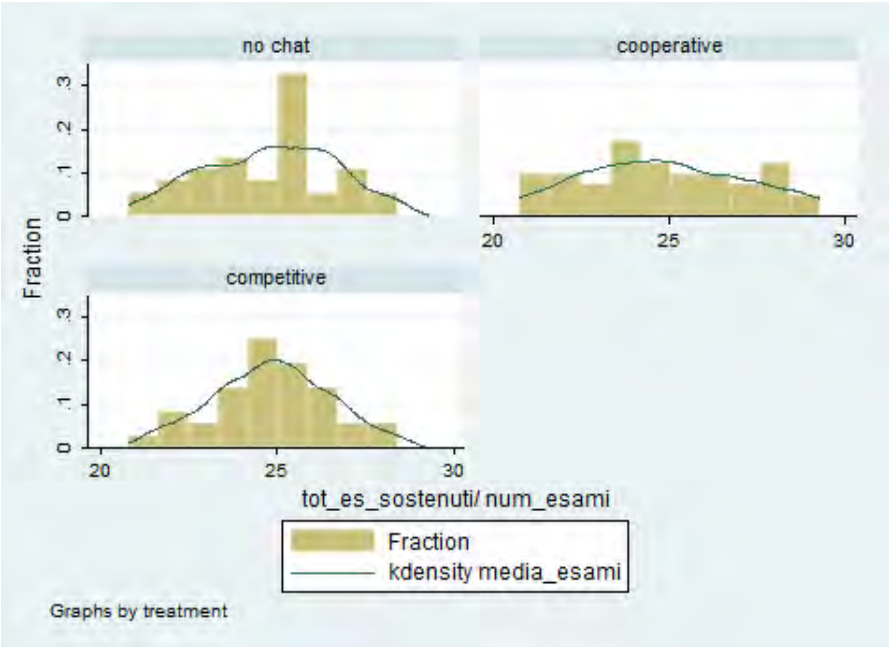
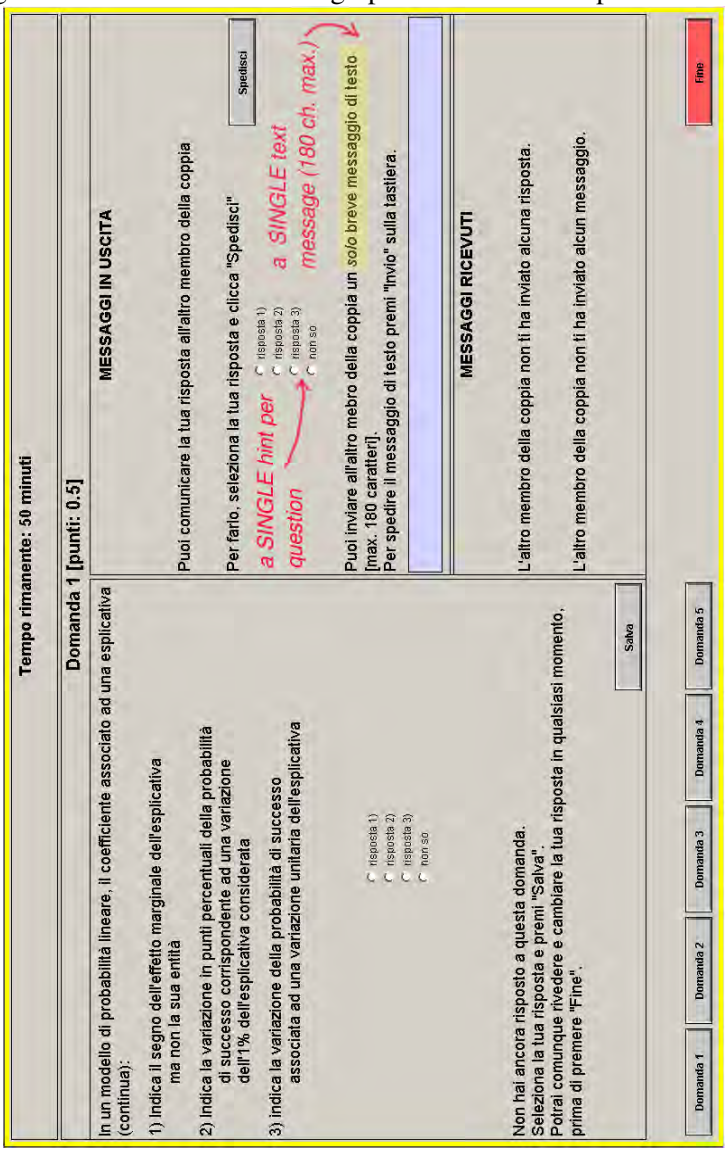


Figure A.4: Screen-shot of the graphical interface for partial exams.



A.4 Examples of chat messages

Figure A.5: Example of use of the chat under the cooperative scheme

A: Come on! Tell me which answers do you need. If you don't get 1.5 points, we will lose the bonus.
B: In my opinion the right one is the 2nd
A: OK! I trust you

Figure A.6: Example of use of the chat under the competitive scheme

A: In this case the 4th is the best answer
B (replies): Why do you pass me this solution? Are you trying to screw me?
C: I know that you are going to pass me the wrong answers.
D: I'm not sure...probably the right answer is the 1st *[she choses the 3rd]*

References

ALMALAUREA, . (2009): "XI Rapporto Almalaurea 2009," Report, Consorzio Interuniversitario AlmaLaurea - MIUR.

ANGRIST, J., D. LANG, AND P. OREOPOULOS (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 1(1), 136–63.

ANGRIST, J., AND V. LAVY (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384–1414.

AZMAT, G., AND N. IRIBERRI (2010): "The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness," Working Papers (Universitat Pompeu Fabra. Departamento de Economía y Empresa), No. 1216.

BLIMPO, M. (2010): "Team Incentives for Education in Developing Countries. A Randomized Field Experiment in Benin," mineo, March 2010 - Stanford University.

BRATTI, M., D. CHECCHI, AND A. FILIPPIN (2008): "Should You Compete or Cooperate with Your Schoolmates?," IZA Discussion Papers 3599, Institute for the Study of Labor (IZA).

BRUNELLO, G., AND R. WINTER-EBMER (2003): "Why Do Students Expect to Stay Longer in College? Evidence from Europe," *Economics Letters*, 80(2), 247–253.

COOPER, D., AND J. KAGEL (2005): "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games," *American Economic Review*, 95(3), 477–509.

DE PAOLA, M., AND V. SCOPPA (2008): "Peer Group Effects in the Academic Performance of Italian Students," *Applied Economics*, pp. 1466–4283.

——— (2010): "Monetary Incentives and Student Achievement in a Depressed Labour Market: Results from a Randomized Experiment," Working Paper n. 06 - 2010, UNICAL, Dipartimento di Economia e Statistica.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2007): "Are Risk Aversion and Impatience Related to Cognitive Ability?," Discussion Paper DP 2735, Institute for the Study of Labor, IZA.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2008): "Representative Trust and Reciprocity: Prevalence and Determinants," *Economic Inquiry*, 46(1), 84–90.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2005): "Individual Risk Attitudes: New Evidence From a Large, Representative, Experimentally-Validated Survey," Discussion Paper DP 1730, Institute for the Study of Labor, IZA.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (forth.): "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences," *Economic Inquiry*.

FALK, A., AND A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–58.

FISCHBACHER, U. (2007a): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10(2), 171–178.

FISCHBACHER, U. (2007b): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

FRYER, R. J. (2010): "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *working paper Harvard University and NBER*.

GNEEZY, U., K. L. LEONARD, AND J. LIST (2009): "Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society," *Econometrica*, 77(5), 1637–1664.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance In Competitive Environments: Gender Differences," *The Quarterly Journal of Economics*, 118(3), 1049–1074.

GNEEZY, U., AND A. RUSTICHINI (2004): "Gender and Competition at a Young Age," *American Economic Review*, 94(2), 377–381.

KREMER, M., E. MIGUEL, AND R. THORNTON (2009): "Incentives to Learn," *Review of Economics and Statistics*, 91(3), 437–456.

LANG, F., D. WEISS, A. STOCKER, AND B. ROSENBLADT (2007): "Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP)," *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 127(1), 183–192.

LAZEAR, E., AND S. ROSEN (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841–64.

LEUVEN, E., H. OOSTERBEEK, AND VAN DER KLAAUW B. (forth.): "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment," ., Journal of the European Economic Association.

LEVITT, S. D., AND J. LIST (2007): "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *Journal of Economic Perspectives*, 21(2), 153–174.

LYLE, D. (2009): "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, 1(4), 69–84.

MANN, H., AND D. WHITNEY (1947): "On a test of whether one of 2 random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, 18, pp. 50–60.

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99(1), 112–145.

NIEDERLE, M., AND L. VESTERLUND (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, 122(3), 1067–1101.

NIEDERLE, M., AND L. VESTERLUND (2010): "Explaining the Gender Gap in Math Test Scores: The Role of Competition," *Journal of Economic Perspectives*, 24(2), 129–44.

PRENDERGAST, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), pp. 7–63.

SACERDOTE, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116(2), 681–704.

STINEBRICKNER, R., AND T. STINEBRICKNER (2008): "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*, 8(1).

WILCOXON, F. (1945): "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, 1(6), pp. 80–83.

ZIMMERMAN, D. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 85(1), 9–23.