Economics of Education Review 30 (2011) 1416–1429

Contents lists available at ScienceDirect

Economics of Education Review

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

Frequency of examinations and student achievement in a randomized experiment

Maria De Paola, Vincenzo Scoppa*

Department of Economics and Statistics, University of Calabria, Italy

ARTICLE INFO

Article history:
Received 10 November 2010
Received in revised form 24 May 2011
Accepted 6 July 2011

JEL classification:
I21
J31
D82

Keywords:
Frequency of examinations
Education production function
Work organization
Feedback provision
Higher education
Randomized evaluation

ABSTRACT

We carry out a randomized experiment involving undergraduate students enrolled at an Italian University attending two introductory economics classes to evaluate the impact on achievement of examination frequency and interim feedback provision. Students in the treated group were allowed to undertake an intermediate exam and were informed about the results obtained, while students in the control group could only take the final exam. The results show that students undertaking the intermediate exam perform better both in terms of the probability of passing the exams and of grades obtained. High ability students appear to benefit more from the treatment. The experiment design allows us to disentangle "workload division or commitment" effects from "feedback provision" effects. We find that the estimated treatment impact is due exclusively to the first effect, while the feedback provision has no positive effect on performance. Finally, the better performance of treated students in targeted examinations seems not to be obtained at the expenses of results earned in other examinations.

1. Introduction

A large and increasing body of economic literature has analyzed educational processes from a theoretical and an empirical point of view, trying to understand the role played by a number of factors, such as class-size, teacher abilities, peer group quality and parents' background in the determination of students outcomes.

Some recent works have tried to investigate the effects produced on student performance by teaching and evaluation practices.[1] An important issue concerning this topic regard the optimal number of examinations, that is, whether it is better to test students more or less frequently, assigning them a smaller or a larger workload. On the one hand, when examinations are frequent and focused on a small number of issues students may find it easier to organize their work, with a positive effect on their learning process. In addition, students used to procrastinate their effort may end up studying more if they have more frequent deadlines. Frequent examinations also offer students interim feedback of their results allowing them to know if their study effort has been appropriate and to become aware of their areas of strength and weakness. In fact, students may need a tangible way to measure their progress during a class. On the other hand, when testing is too recurrent, students may not have enough time to deepen their knowledge and to understand the relationships among the range of concepts covered in a given subject. Moreover, they may be exposed to an excessive amount of stress.

The effects of test frequency and feedback provision on student achievement have been mainly investigated from a pedagogical and psychological point of view. Tuckman

* Corresponding author.
E-mail addresses: m.depaola@unical.it (M. De Paola), v.scoppa@unical.it (V. Scoppa).

[1] See, for example, Schwerdt and Wuppermann (2011) and Schacter and Thum (2004).

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.07.009

(1997), Deck (1998), Hattie and Timperley (2007), among others, show that more frequent testing produce positive effects on student performance. Ariely and Wertenbroch (2002) analyze the use of deadlines as a commitment device. According to their analysis, based on a sample of students enrolled at a MIT class, people have self-control problems and, to avoid them, choose costly deadlines that seem to improve their academic performance.

The economic literature has instead scarcely investigated this issue. Most of the existing literature examines feedback provision as an organizational design problem. In relative performance evaluations, for example, organizations decide whether to inform or not their employees about their relative standing at intermediate stages of the competition. In principle, information provision has ambiguous effects: performance might improve when workers who are obtaining a bad performance decide to work harder and try to avoid failure and when workers who are doing well become even more enthusiastic; on the other hand, performance tends to worsen when informed underdog become discouraged and workers getting a good performance, knowing that they are well ahead of the other colleagues, decide to shirk. These aspects have been analyzed from a theoretical point of view by a number of recent papers showing that feedback on past performance can affect current performance either directly – if past and current performances are substitutes or complements in the agent's utility function – or indirectly, by revealing information on the marginal return to current effort (Aoyagi, 2010; Ederer, 2010; Lizzeri et al., 2002; Perry & Reny, 1999; Yildirim, 2005).

Empirical investigations trying to shed light on these effects are scant and lead to ambiguous results. Eriksson, Poulsen, and Villeval (2009) show, through a laboratory experiment, that under a piece-rate pay scheme feedback on relative performance does not improve performance, but under the tournament pay scheme there are positive peer effects. More precisely, under the latter scheme underdogs do not give up even in case of large gaps with their competitors and workers doing well do not slack off. In a similar vein, Freeman and Gelber (2010) find that individual performance improves when tournament participants are informed about their own and their competitors' past results.

Some works have tried to investigate these issues focusing on educational contexts. Bandiera, Larcinese, and Rasul (2009) study the effect of providing University students with interim feedback information regarding their own performance. The authors show a positive and statistically significant effect of feedback provision on student final performance. In a similar vein, Azmat and Iriberri (2010), considering high school students in Spain, investigate the effect of informing students on whether they were performing above or below the class average. It emerges that students receiving this type of information obtain better grades.

In this paper we contribute to this emerging literature analyzing the effect of test frequency and feedback provision on a sample of Italian University students. Currently, in Italy University classes are typically organized in long modules offering about 60–80 h of teaching activities. At the end of each class students take an exam and usually no intermediate assessment of student knowledge is undertaken. The current organization is the result of the introduction of some new University rules, which have imposed a limit to the maximum number of exams needed to gain a First Level Degree. Before this change, teaching activity was organized in short modules (30–40 h of teaching activity) and assessment was undertaken at the end of each module. The limit of a maximum number of exams introduced by the new law has forced Universities to reorganize their academic curricula and to unify short modules. The effects of this change have not been investigated yet. Nevertheless, if frequent examinations encourage students to study more and allow them to receive useful feedback then perhaps it would be worthwhile to put additional effort in trying to identify the optimal number of exams and to improve teaching organization.

To investigate the effects of test frequency and interim feedback on student performance we have conducted a randomized field experiment involving 344 undergraduate students enrolled at a middle sized Italian public University and attending two introductory economics classes. Students participating at the experiment were randomly assigned to a control group and to a treatment group. Students in the treatment group were allowed to undertake an intermediate examination covering the first part of the course material and a final exam covering exclusively the second part, while students in the control group were permitted to undertake exclusively the final examination (covering the whole course material) at the end of the course program (as established by the University rules).

We have decided to adopt this experimental design instead of one considering midterm exams but imposing students a final exam covering the whole course material because in Italy midterms of this kind are scarcely diffused and a similar design would not be related to the reform we aimed to evaluate. More importantly, using that design we would have been able to investigate exclusively feedback effects, while we were also interested in the workload and procrastination effects.

Thus, differently from Bandiera et al. (2009) and Azmat and Iriberri (2010) who only focus on the effect of feedback provision on student performance, we couple feedback information provision with workload division and try to disentangle their relative importance.

We firstly investigate "intention-to-treat" effects, considering as treated all the students assigned to the treatment group by the random procedure. Our results show that students in the treatment group perform significantly better than students in the control group in terms of probability of passing exams and of grades obtained. Subsequently, we investigate the impact of the effective participation to the treatment, using as an instrument the random assignment to the groups: adjusting the "intention-to-treat" effect for non-compliance leads to a stronger and highly significant effect of intermediate examination on student performance.

As the treatment effect can be seen as the combined result of a "feedback provision" effect and a "workload division or commitment" effect, we propose a framework to disentangle these effects. We find that the uncovered impact is due exclusively to the "workload division or commitment" effect, while the feedback provision has no positive impact on performance.

Finally, examining students' performance in other exams, we show that the improvement detected in the performance of treated students is not the result of a substitution effect. As a matter of fact, treated students did not obtain a worse performance in non targeted examinations compared to students in the control group.

The paper is organized as follows. In Section 2 the design of the experiment is explained and some information on the Italian University system are provided. Section 3 shows the effects of examination frequency and feedback provision on the probability of passing exams and on grades obtained by students. In Section 4 we investigate the treatment effects on students who have effectively undertaken the intermediate examination using an instrumental variable strategy. In Section 5 we analyze the existence of heterogeneous effects. In Section 6 we disentangle feedback effects from workload division or commitment effects. Section 7 investigates the existence of substitution effects. Section 8 concludes.

2. Experiment description and data

The experiment we conducted has involved 344 students enrolled at the Microeconomics and Macroeconomics classes offered by the First Level Degree Course in Business and Administration (BA hereafter) at the University of Calabria in the academic year 2009–2010. The University of Calabria is a middle-sized public University located in the South of Italy. It has currently about 34,000 students enrolled in different Degree Courses and at different levels of the Italian University system.

Since the 2001 reform, the Italian University system is organized around three main levels: First Level Degrees (3 years of legal duration), Second Level Degrees (2 years more) and Ph.D. degrees. In order to gain a First Level Degree students have to acquire a total of 180 credits. Students who have acquired a First Level Degree can undertake a Second Level Degree (acquiring 120 more credits). After having accomplished their Second Level Degree, students can enroll in a Ph.D. degree. The 2001 reform has introduced a credit system aimed to facilitate mutual recognition of degrees among European countries and has given autonomy to Universities. Among the unintended consequences of the reform there has been a proliferation of short classes. For example, at the BA Degree offered by the University of Calabria, a typical class allowed to acquire 5 credits, and students had to undertake more than 30 exams in order to acquire 180 credits. The introductory classes in Mathematics, Microeconomics, Macroeconomics, Statistics etc., were divided in two modules (5 credits each) assessed through two different exams at the end of each teaching period. In order to limit the proliferation of classes, a new rule, known as "DM 270", coming into effect in the academic year 2008/09, has imposed a maximum of 20 exams for First Level Degrees leaving unchanged the number of credits to acquire (180).

At the University of Calabria the "DM 270" has led to a reorganization of the academic curriculum in order to reduce the number of exams students had to undertake. In particular, some classes, such as Macroeconomics and Microeconomics, initially split in two modules, were unified in a single class (10 credits) and students were required to undertake a unique exam.

One of the ideas behind the "DM 270" was to give students more time to prepare for exams and avoid them the stress deriving from sitting to a large number of examinations. However, as explained above, reducing the number of exams might also produce some negative effects on student performance.

By comparing the average performance of students enrolled at the BA Degree, immediately before and after the introduction of "DM 270", it emerges a drop in the number of credits acquired by students. Students enrolled in 2007/2008, under the old regime, have acquired 38.9 credits during their first year, while their counterparts enrolled in 2008/2009, under the new system, have acquired only 32.4 credits. However, these figures might be driven by temporal trends and unobserved changes in student and instructors' characteristics. Then, in order to try to shed light on this issue, we have decided to undertake the experiment described in this paper.

At the beginning of the Microeconomics and Macroeconomics classes (in March 2010) students were informed of the experiment both through presentations during the teaching hours and through a letter, sent to all students, explaining the format of the experiment.

We asked students to register for joining the experiment. We did not consider in the experiment the small fraction (around 10%) of students who did not register.[2] On the basis of the available administrative information on students' characteristics, we proceeded to the stratification of students participating to the experiment according to the following variables: class attended (Microeconomics or Macroeconomics); gender; type of High School attended (3 categories: Lyceum; Technical Schools; Vocational and other types of schools); final grade obtained at High School (split in 4 categories corresponding to quartiles).

Following this procedure the 344 students were allocated to 48 non null groups. Within each group, one half of students was randomly assigned to the treatment group – allowed to take the intermediate exam – and the other half was assigned to the control group, which could take a unique exam at the end of the class, without an intermediate exam.[3] We ended up with 172 students assigned to the treatment group, and 172 to the control group.

The random assignment procedure was carried out at the presence of students. They were also informed by e-mail of their assignment status and the list of students belonging to the treatment and control group was published on the classes' web-pages.

The Microeconomics and Macroeconomics classes were taught to students enrolled, respectively, at the first and

[2] These students may be low motivated students, not attending classes and not particularly involved with studying activities.

[3] When the number of students included in each stratified group was not even, one student was first assigned randomly to the treatment or control group.

second year of the BA Degree. Both these classes started in March 2010 and lasted until June. Each class program consisted in 60 h of teaching activity and 20 h of laboratory. Treatment and control groups attended the class in the same room, at the same time and with the same instructor and teaching material.

After the first 30 h of teaching activity (and 10 h of laboratory) there was a break of two weeks (May 2010). The classes teaching programs were then naturally divided in two parts (that is, Microeconomics-1 and Microeconomics-2 and Macroeconomics-1 and Macroeconomics-2). During the break period, students in the treatment group were allowed to undertake an intermediate exam covering respectively the Microeconomics-1 and the Macroeconomics-1 program. The intermediate exam consisted in 30 multiple choice questions and lasted 1 h. 77% of students in the treatment group have effectively undertaken the test.

All the students were informed about the results obtained at the test and the correct answers to the interim exam have been given both to treated and control students. Moreover, examples of the questions proposed at the exam were presented during laboratory hours. Therefore, we are confident that treated students, participating at the interim exam, have not collected more information on the type of questions proposed at the exams.

Teaching activities re-started in the mid of May and lasted for other 30 h (+10 h of laboratory), until the end of June. In July students in the treatment group were required to complete their examination by taking an exam covering exclusively the second part of the teaching program (respectively, Microeconomics-2 and Macroeconomics-2), consisting in 30 multiple choice questions (in 1 h). The final grade obtained by these students was given by the average of the grades earned at the first and at the second part of the exam.

Students in the control group have, instead, undertaken the exam covering the whole course program (the first plus the second part). This exam was held in July and has consisted in 30 + 30 multiple choice questions. These students had the double of time allowed to students in the treatment groups to complete the entire examination (2 h broken by an interval of 30 min).

As regard the first module, since each exam question was randomly selected from a large test-bank, the exam for treated students can be considered equivalent to the exam for students in the control group. The second module was identical for treated and control students.

We had a number of non-complier students: 22 treated students never shown (neither to the intermediate nor to the whole examination); 43 treated students shifted to the control group (17 did not undertake the intermediate exam; while 26 chose to undertake the entire exam after having participated at the intermediate exam).[4] Finally, 27 students took the intermediate exam, but did not complete the exam undertaking the second part.

Table 1
Descriptive statistics.

Variables	Obs	Mean	Std. dev.	Min	Max
Pass	344	0.348	0.477	0	1
Grade	236	15.371	8.571	1	31
Treatment (intermediate exam)	344	0.500	0.501	0	1
Macroeconomics	344	0.439	0.497	0	1
Female	344	0.613	0.488	0	1
Credits	344	14.245	6.865	0	28.667
High School Grade	344	88.828	8.479	68	100
Technical Schools	344	0.453	0.499	0	1
Lyceum	344	0.416	0.494	0	1
Late Enrollment	344	0.140	0.347	0	1
Resident near University	344	0.558	0.497	0	1

Notes: Grades in each class range from 18 to "30 cum laude" (set equal to 31). High School Grade ranges from 60 to 100.

In this first part of our analysis we focus on "intention-to-treat" effects, considering as treated all the 172 students randomly selected. In Sections 4 and 5 we proceed both by using an instrumental variable strategy and by defining a more restricted sample of students.

We measure student performance considering both the probability of passing the target examinations and the grades obtained. In the Italian system, passing grades range from 18 to "30 cum laude", which we consider equal to 31. We observe both grades in passed examinations (18–31) and in failed examinations. Only exams undertaken until the 31st of July 2010 were taken into account in determining student performance.

Table 1 provides descriptive statistics for our sample of students. About 61% of students were female. 44% of students attended the Macroeconomics class while 56% attended Microeconomics. High School Grade ranged from 60 (the minimum passing grade) to 100 (the maximum grade), with a mean of 88.8. Students mainly came from Technical Schools (45%) and Lyceums (about 42%). As an additional measure of student ability we have considered the number of credits acquired by students until the beginning of the experiment, *Credits*, and to make comparable the two cohorts of students involved in the experiment we have divided the number of credits by the number of exam sessions they had available (1 for first year students and 3 for second year students). The average number of credits acquired in a semester was 14.24. About 55% of students came from the same province (NUTS-3) where the University is located (*Resident near University*). 14% of sample students did not enroll at University immediately after High School graduation, but a year or more later (*Late Enrollment*).

At the end of exam session, 35% of students passed the exams considered in this experiment. The average grade (ranging from 1 to 31 is 15.4). The average grade for students passing the exams was 22.6.

In the first two columns of Table 2 are reported, by treatment groups, means for a number of individual characteristics. In the third column we report differences in

[4] We were forced by the University rules to allow students willing to undertake the entire exam after having participated at the intermediate exam to do so (these are students who obtained a very low grade at the intermediate exam).

Table 2
Student characteristics across treatment and control groups.

	Means		Differences (s.e.)
	Treatment	Control	Treatment vs. control
Macroeconomics	0.442	0.436	0.006 (0.054)
Female	0.610	0.616	−0.006 (0.053)
Credits	14.031	14.459	−0.428 (0.741)
High School Grade	88.564	89.093	−0.529 (0.915)
Technical Schools	0.453	0.453	0.000 (0.054)
Lyceum	0.419	0.413	0.006 (0.053)
Late Enrollment	0.157	0.122	0.035 (0.037)
Resident near University	0.552	0.564	−0.012 (0.054)

Notes: Standard errors are reported in parentheses.

means between treatment and control groups (standard errors are reported in parentheses). Results show that the randomization has been successful in creating comparable treatment and control groups along the observable characteristics: there are no significant differences by treatment status in class, gender, number of credits acquired, High School Grade, type of High School attended, Late Enrollment and place of residence.

3. Treatment effects on student achievement

In this section we analyze the effects of treatment both on the student probability of passing examinations and on grades obtained. We focus on "intention-to-treat" effects considering as treated all students randomly assigned to the treated group.

3.1. The effect of treatment on the probability of passing the exam

We use a Probit model to estimate the probability of passing the examination targeted in the experiment (either Microeconomics or Macroeconomics):

$$\Pr(Pass_i = 1 | Treatment_i, X_i, \mu_i) = \Phi(\beta_0 + \beta_1(Treatment_i) + \phi X_i + \mu_i) \quad (1)$$

Our dependent variable $Pass_i$ takes the value of one if student i passed the exam, and 0 otherwise, $Treatment_i$ is a dummy variable which takes value of 1 if student i was assigned to the treatment group while takes value of 0 if i belong to the control group, X_i is a vector of the individual characteristics of i (measures of his/her ability and personal characteristics), μ_i is a dummy for the Macroeconomics class, ε_i is an error term capturing idiosyncratic shocks or unobserved student characteristics.

For each student we observe separately the outcome of the two modules composing each exam. However, in this section we have organized data at the student level taking as a measure of student performance the average outcome at the two modules. This implies that we consider the exam as passed when the average grade obtained by the student is equal or higher than 18.

The coefficient on *Treatment* captures the "intention-to treat" effect on student performance (with respect to the control group). The marginal effects from Probit estimates are reported in Table 3. In all the specifications, standard errors are robust to heteroskedasticity.

In column 1 we regress $Pass_i$ on the *Treatment* and *Macroeconomics* dummies, without other controls. It emerges that students in the treatment group have a probability of about 15 percentage points higher to pass the exam. The coefficient is significant at the 1% level. In column 2 we control for individual characteristics: the probability of passing the exam increases by 20 percentage points for students in the treatment group.

In columns 3 and 4 we run separate regressions for the Macroeconomics and Microeconomics classes. The effect appears relevant for both classes. We have checked whether the difference in the impact for the two classes is statistically significant estimating specification 2 including an interaction term between *Treatment* and *Macroeconomics* (not reported to avoid to clutter the table). It emerges that the treatment effect is significantly higher for students attending the Macroeconomics class. Probably this is due to the fact that students find it harder to pass the Macroeconomics exam and the benefits of the workload division or of the feedback provision, allowed by the treatment, may be higher for tougher exams.

Columns 5 and 6 show results for males and females separately: it seems that males react more than females to the treatment, but when we use an interaction term to test the statistical significance of this difference we are not able to reject the null hypothesis (results not reported).

As robustness checks we have also used a linear probability model instead of a Probit. The OLS estimates are very similar to the Probit estimates presented in Table 3 and are not reported.

The effects of control variables are consistent with the findings emerging from the education literature. Both our measures of student predetermined ability, *High School Grade* and *Credits*, produce a positive and statistically significant effect on academic performance. Once we control for these measures of abilities neither the dummy *Lyceum* nor the dummy *Technical Schools* produce a significant effect. The dummy *Female* is not statistically significant. The probability of passing the Macroeconomics examination turns out to be lower than the probability of passing Microeconomics.

We are confident that changes in student behavior deriving from the fact that they are aware of being observed (the so-called Hawthorne effect) are not particularly relevant in our case. Firstly, we did not inform students that the experiment was aimed at evaluating the recent reform, Secondly, students, especially freshmen, are not well informed about the changes over time of the University rules. Finally, even in the case in which they had realized that the experiment was related to the reform, they had no reason to believe that the experiment results may produce any change in the already implemented rules (unfortunately, policy evaluations are rare in Italy and, even

Table 3
Estimates of the treatment effect on the probability of passing the exam. Dependent variable: *Pass*.

	(1) All	(2) All	(3) Macro	(4) Micro	(5) Females	(6) Males
Treatment (intermediate exam)	0.153***	0.205***	0.290***	0.136*	0.161**	0.262***
	(0.051)	(0.054)	(0.077)	(0.074)	(0.072)	(0.083)
Macroeconomics	−0.071	−0.187***			−0.198**	−0.176**
	(0.051)	(0.057)			(0.077)	(0.085)
Female		−0.042	−0.041	−0.041		
		(0.057)	(0.086)	(0.078)		
High School Grade		0.013***	0.011***	0.015***	0.012***	0.012**
		(0.004)	(0.005)	(0.005)	(0.004)	(0.006)
Credits		0.033***	0.028***	0.036***	0.040***	0.028***
		(0.005)	(0.009)	(0.006)	(0.007)	(0.008)
Technical Schools		0.005	0.029	−0.002	0.210*	−0.241*
		(0.091)	(0.130)	(0.130)	(0.127)	(0.131)
Lyceum		0.112	0.162	0.089	0.270**	−0.127
		(0.093)	(0.132)	(0.133)	(0.133)	(0.129)
Late Enrollment		0.021	0.089	−0.013	0.018	0.056
		(0.088)	(0.164)	(0.103)	(0.131)	(0.131)
Resident near University		0.058	0.053	0.068	−0.057	0.231***
		(0.056)	(0.081)	(0.075)	(0.070)	(0.089)
Observations	344	344	151	193	211	133
Pseudo *R*-squared	0.024	0.244	0.256	0.247	0.347	0.168
Log-likelihood	−217.180	−168.159	−69.703	−96.345	−88.383	−72.322

Notes: The table reports marginal effects from Probit estimates evaluated at the mean values of the explanatory variables in the sample. Robust standard errors are reported in parentheses.
* Coefficients are statistically significant, 10% level.
** Coefficients are statistically significant, 5% level.
*** Coefficients are statistically significant, at 1% level.

when undertaken, they often do not produce direct effects on policy-maker decisions). However, as in all the experiments in which the treatment cannot be blind (see Duflo, Glennerster, & Kremer, 2007), it is not possible to exclude that being assigned to the treatment group has determined some sort of "empowerment effect" in the students concerned leading them to provide more effort.

3.2. *The effect of treatment on grades*

In this section we estimate the effect of the treatment on student grades. Our dependent variable is the grade obtained by student i at the exam. To take into account the fact that *Grade* is censored, since a number of students did not sit for the exam and we do not have information on their performance, we use a Tobit model expressing the observed outcome $Grade_i$ in terms of a latent variable $Grade_i^*$:

$$\begin{aligned} Grade_i^* &= \beta_0 + \beta_1(Treatment_i) + \phi X_i + \mu_i + \varepsilon_i \\ Grade_i &= \max(0, Grade_i^*) \end{aligned} \tag{2}$$

We set equal to zero the variable *Grade* for absent students.

Table 4 reports the Tobit estimates using the same specifications as in Table 3 (the lower limit has been set equal to 0 since the minimum grade observed is 1). In Panel A are reported the marginal effects of explanatory variables on the expected grade conditional on being uncensored, that is, given that the student has sit for the exam. In all the specifications it emerges that taking an intermediate exam has a positive and statistically significant impact on student achievement: the expected grade increases by about 5 points when controlling for individual characteristics.

On the basis of the Tobit estimations, we have also determined the marginal effect of treatment on the probability that a student will undertake the exam (that is, the effect on the probability that the grade is greater than 0). These effects are reported in Panel B of Table 4. Undertaking the intermediate examination increases the probability of sitting for the exam by about 25 percentage points.

Treatment effects do not differ much according to the type of class attended by the students (columns 3 and 4) and according to gender (columns 5 and 6).[5]

4. Treatment effects on students effectively undertaking the intermediate exam: IV estimates

In the previous section we have analyzed "intention-to-treat" effects, since we have considered all students in the treatment group independently on their effective participation at the intermediate exam. Therefore, the estimated effects are diluted by the fact that some treated students may actually not have undertaken the intermediate exam (see Angrist & Pischke, 2009; Bloom, 1984).

As a matter of fact, students in the treatment group may refuse treatment choosing to not sit for the intermediate exam or choosing after the intermediate exam to retake the entire exam. On the other hand, the 172 students assigned to the control group could not shift to the treatment group.

In this section we analyze the impact of the effective treatment defining three alternative treated groups.

[5] We have also estimated a Tobit model with left and right censoring using 31 as the upper limit for right censoring. Results (not reported) are very similar.

Table 4
Tobit estimates of the treatment effect on grades.

	(1) All	(2) All	(3) Macro	(4) Micro	(5) Females	(6) Males
Panel A. Marginal effects: conditional on being uncensored						
Treatment (intermediate exam)	4.489***	4.993***	5.124***	4.839***	4.890***	4.901***
	(0.782)	(0.702)	(1.023)	(0.966)	(0.856)	(1.211)
Macroeconomics	−1.057	−2.577***			−2.920***	−1.875
	(0.777)	(0.712)			(0.874)	(1.229)
Female		−0.215	−0.593	0.259		
		(0.709)	(1.075)	(0.957)		
High School Grade		0.088*	0.019	0.145**	0.045	0.165**
		(0.045)	(0.071)	(0.061)	(0.055)	(0.081)
Credits		0.528***	0.510***	0.542***	0.593***	0.466***
		(0.057)	(0.091)	(0.074)	(0.068)	(0.109)
Technical Schools		−0.808	−0.344	−1.247	0.439	−3.083
		(1.115)	(1.563)	(1.585)	(1.396)	(1.913)
Lyceum		1.075	1.565	0.697	1.835	−0.516
		(1.114)	(1.558)	(1.585)	(1.403)	(1.884)
Late Enrollment		0.863	0.711	0.833	1.046	0.742
		(0.987)	(1.536)	(1.315)	(1.247)	(1.616)
Resident near University		0.144	−0.008	0.195	−0.948	1.859
		(0.708)	(1.041)	(0.975)	(0.865)	(1.285)
Panel B. Marginal effects: probability uncensored						
Treatment (intermediate exam)	0.224***	0.257***	0.296***	0.226***	0.253***	0.250***
	(0.040)	(0.037)	(0.061)	(0.046)	(0.045)	(0.063)
Macroeconomics	−0.054	−0.139***			−0.157***	−0.100
	(0.039)	(0.037)			(0.046)	(0.064)
Female		−0.011	−0.034	0.012		
		(0.037)	(0.064)	(0.045)		
High School Grade		0.005*	0.001	0.007**	0.002	0.009**
		(0.002)	(0.004)	(0.003)	(0.003)	(0.004)
Credits		0.028***	0.030***	0.026***	0.031***	0.024***
		(0.003)	(0.005)	(0.004)	(0.004)	(0.006)
Technical Schools		−0.043	−0.020	−0.060	0.023	−0.159
		(0.058)	(0.092)	(0.075)	(0.074)	(0.099)
Lyceum		0.055	0.089	0.033	0.094	−0.027
		(0.058)	(0.092)	(0.075)	(0.074)	(0.097)
Late Enrollment		0.042	0.040	0.037	0.051	0.037
		(0.052)	(0.091)	(0.062)	(0.066)	(0.084)
Resident near University		0.008	−0.000	0.009	−0.050	0.099
		(0.037)	(0.062)	(0.046)	(0.046)	(0.066)
Observations	344	344	151	193	211	133
Pseudo R-squared	0.017	0.069	0.068	0.072	0.091	0.048
Log-likelihood	−1015	−960.8	−410.8	−548.5	−578.8	−376.7

Notes: The table reports marginal effects from Tobit estimates. Robust standard errors are reported in parentheses.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

Firstly, we consider as treated the 133 students (out of 172) assigned to the treatment group who have effectively undertaken the intermediate exam regardless of whether they have subsequently shifted to the entire exam (“Effective Treatment 1”). Secondly, we define a more restricted treated group excluding from the 133 students undertaking the intermediate exam 26 students who have decided to take in July the entire exam (“Effective Treatment 2”). Finally, we use “Effective Treatment 2”, but we exclude from the control group all those students who did not sit for the exam (“Effective Treatment 3”).

As expected, non-compliers, that is, subjects dropping out or switching from the treatment to the control group, have different characteristics from treated students who comply with the treatment. More in detail, the 43 treated subjects who shifted to the control group are characterized by a lower High School Grade and by a higher probability of not enrolling regularly. These differences are statistically significant at the 10% level, while differences along other dimensions are not statistically significant.[6] Secondly, the 27 treated students who have undertaken the intermediate exam and did not show to the final exam are characterized by a lower *High School Grade* and have acquired a lower number of credits compared to those who did show at the second exam. Finally, attrition is not the same across treated and control groups: treated students who did not

[6] 26 out of the 43 “shifters” are students who were assigned to the treatment group and did not undertake the intermediate exam but took part in the final exam. The achievement of these students in terms of grade (14.12) and probability of passing the exam (0.35) is much worse in comparison to other treated students (grade = 19.8 and pass = 0.75). Similarly, their pre-determined characteristics tend to be worse (for example, *High School Grade* is 87 instead of 90).

sit the exam are 22, while drop out students in the control group are 86.[7]

The differences existing between compliers and non-compliers confirm a self-selection problem: to deal with the endogeneity problems related to the choice of the effective treatment, we adopt, following the literature, an instrumental variable estimation strategy, using as an instrument for the effective participation the randomly assigned treatment status (*Treatment*).

In Table 5 are reported the Instrumental Variable estimates. Form first stage regressions (Panel B), as expected, it emerges that the assigned treatment status is highly significant in predicting the effective treatment. In Panel A are reported second stage results. In columns 1, 2 and 3 we consider as dependent variable *Pass*, reporting the marginal effects of an IV-Probit model. In column 1 we focus on "Effective Treatment 1", in column 2 with "Effective Treatment 2", while in column 3 we deal with "Effective Treatment 3". In columns 4, 5 and 6 the dependent variable is *Grade*. We report IV-Tobit estimates (the marginal effects conditional on being uncensored) considering respectively "Effective Treatment 1", "Effective Treatment 2" and "Effective Treatment 3".

Results show that adjusting "intention-to-treat" effects for non-participation leads to a stronger impact of the treatment on student performance measured both as the probability of passing the exam and by the grade obtained. The increase in the effect is proportional to the reciprocal of the participation rate.

When we exclude from the sample students that did not sit for the exam, we find smaller treatment effects, since one of the effects of treatment was that of inducing students to take the exam regularly, while a large part of students in the control group have dropped out. However, we find a positive treatment effect even in the sample excluding drop-out students.

5. Heterogeneous effects across students with different abilities

In this section we investigate whether the effects of treatment differ according to students' abilities. To split students into a high ability and a low ability group, we have used three different measures of their skills: (1) the grades obtained at high school (*High School Grade*); (2) the number of credits gained by the beginning of the experiment (*Credits*); (3) their predicted performance (*Composite Ability*).

To obtain the latter composite measure of student ability, we follow Angrist and Lavy (2009) and we first estimate a model for student performance considering exclusively students in the control group, using as explanatory variables individual characteristics such as gender, type of high school attended, High School Grade. The estimated coefficients of this model are then used for predicting the performance of students in treated and control groups, on the basis of their effective characteristics.

[7] These statistics are not reported and available upon request.

In Table 6 are reported a number of Probit estimates showing the effects of treatment on the probability of passing exams respectively on high and the low ability students. In column 1 we measure student ability considering *Credits* and to investigate whether there are heterogeneous effects we use the interaction variable *Credits* × *Treatment*, where *Credits* is demeaned. The coefficient on the interaction term is positive and statistically significant at the 10% level, suggesting that the positive impact of more frequent exams increases with student ability.

To better describe this aspect we run separate regressions for students with ability above and below the median according to the three measures of ability we have available. In columns 2 and 3 students are grouped according to the number of *Credits*, in columns 4 and 5 using the *High School Grade* and in columns 6 and 7 we define the High and Low ability groups on the basis of their predicted performance, *Composite Ability*. Changing the measure of student ability does not change the results: the effect of treatment on high ability students is always stronger, while the effect for low ability students is not significantly different from zero (except in the case of *Composite Ability*).

The same findings emerge also when we use as dependent variable the grades obtained at exams (estimates are not reported and available upon request).

In Table 7 we replicate the specifications reported in Table 6 but we estimate the impact of the effective treatment using an IV estimation strategy to handle endogeneity problems. We show the effects obtained when considering "*Effective Treatment 2*". To save space, first stage results are not reported. We find that the effective treatment effects does not vary much with abilities. Moreover, while the intention to treat effect for low ability student is typically not significant, the effective treatment effect is positive also for these students.

6. Disentangling feedback and workload division effects

The positive impact of treatment on student performance emerging in previous estimates can be interpreted as the joint result of a feedback provision effect and the effects deriving from the workload division and commitment obtained defining more frequent deadlines.

In this section we propose a framework to try to disentangle the feedback provision effect from the other two effects. At this aim, instead of considering student performance on the whole examination, we have organized our data at student-module level to have information on the results obtained by students separately for the two modules of each examination (Microeconomics-1 and Microeconomics-2; Macroeconomics-1 and Macroeconomics-2).

We exploit the fact that treated students undertaking module 1 compared to control students undertaking the same module benefit exclusively from a workload division or commitment effect, while no feedback effect is at work. Then, comparing the performance of treated and control

Table 5
IV estimates of the effects of effective participation to the treatment.

	(1) Probit	(2)	(3)	(4) Tobit	(5)	(6)
	Pass	Pass	Pass	Grade	Grade	Grade
Effective Treatment 1	0.264***			6.446***		
	(0.079)			(0.991)		
Effective Treatment 2		0.341***			8.703***	
		(0.098)			(1.364)	
Effective Treatment 3			0.194*			1.848*
			(0.117)			(1.211)
Macroeconomics	0.181***	0.172***	0.199***	−2.334***	−2.181***	−2.783***
	(0.058)	(0.059)	(0.084)	(0.687)	(0.705)	(0.940)
Female	−0.053	−0.066	−0.069	−0.602	−0.613	−0.946
	(0.059)	(0.060)	(0.077)	(0.702)	(0.714)	(0.899)
High School Grade	0.013***	0.012***	0.019***	0.088**	0.070*	0.227***
	(0.004)	(0.004)	(0.005)	(0.043)	(0.043)	(0.055)
Credits	0.035***	0.034***	0.040***	0.554***	0.529***	0.741***
	(0.005)	(0.005)	(0.008)	(0.056)	(0.056)	(0.081)
Technical Schools	−0.017	−0.031	−0.005	−0.944	−1.346	−0.603
	(0.091)	(0.090)	(0.120)	(1.042)	(1.054)	(1.378)
Lyceum	0.106	0.095	0.097	1.121	0.819	1.534
	(0.092)	(0.092)	(0.113)	(1.049)	(1.068)	(1.282)
Late Enrollment	0.023	0.033	0.027	0.818	1.249	1.084
	(0.088)	(0.088)	(0.108)	(1.014)	(1.060)	(1.182)
Resident near University	0.047	0.048	0.118	−0.070	−0.153	0.891
	(0.056)	(0.056)	(0.077)	(0.694)	(0.703)	(0.892)
Observations	344	344	236	344	344	236
Panel B: first stage						
Treatment (randomly assigned)	0.773***	0.626***	0.718***	0.773***	0.626***	0.718***
	(0.0317)	(0.036)	(0.037)	(0.0317)	(0.036)	(0.037)
Macroeconomics	−0.044	−0.052	−0.055	−0.044	−0.052	−0.055
	(0.032)	(0.037)	(0.047)	(0.032)	(0.037)	(0.047)
Female	−0.066*	−0.049	−0.042	−0.066*	−0.049	−0.042
	(0.034)	(0.039)	(0.049)	(0.034)	(0.039)	(0.049)
High School Grade	−0.000	−0.002	0.005	−0.000	−0.002	0.005
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Credits	−0.001	−0.001	−0.002	−0.001	−0.001	−0.002
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
Technical Schools	0.036	0.068	0.037	0.036	0.068	0.037
	(0.053)	(0.060)	(0.077)	(0.053)	(0.060)	(0.077)
Lyceum	0.007	0.036	0.007	0.007	0.036	0.007
	(0.053)	(0.071)	(0.077)	(0.053)	(0.071)	(0.077)
Late Enrollment	0.028	−0.036	−0.088	0.028	−0.036	−0.088
	(0.048)	(0.058)	(0.075)	(0.048)	(0.058)	(0.075)
Resident near University	0.028	0.038	0.058	0.028	0.038	0.058
	(0.035)	(0.039)	(0.039)	(0.035)	(0.039)	(0.039)
First-stage *F*-statistics (*p*-value)	594.92 (0.000)	294.93 (0.000)	353.34 (0.000)	594.92 (0.000)	294.93 (0.000)	353.34 (0.000)

Note: In the first stage the dummy for the assigned treatment is used as an instrument for "Effective Treatment 1", "Effective Treatment 2" and "Effective Treatment 3". Standard errors (reported in parentheses) are corrected for heteroskedasticity.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

students at this module we are able to disentangle the workload/commitment effect. Instead, to isolate the feedback effect, we compare the difference in performance of treated students between the two modules. Treated students in both modules benefit of workload/commitment effects, but only for the second module they can obtain positive effects from feedback. Then, the difference in the performance obtained by these students at the two modules should reflect feedback provision, given that we are able to neutralize any eventual heterogeneity in the difficulty level of the first and second module by subtracting the difference in performance at the two modules obtained by control students. Unfortunately, we are not able to distinguish the workload division effect from the commitment effect deriving from frequent deadlines since they overlap in each module.[8]

Formally, we assume that student performance is determined as follows:

$$Y_{ij} = \alpha_0 + \alpha_F F_{ij} + \alpha_W W_i + \alpha_D D_2 + \phi X_i + \varepsilon_{ij} \quad (3)$$

where Y_{ij} is the performance of student i at module j, with $j = 1, 2$. The performance is affected by feedback provision,

[8] In order to distinguish between the effects of commitment and workload division, it would be necessary to observe the distribution of study effort along time (as done, for example, by Burger, Charness, & Lynham, 2011) to verify if student effort is more uniform when they are assigned to the treatment group.

Table 6
Probit estimates. Heterogeneous effects according to student abilities.

	(1) All	(2) High ab.	(3) Low ab.	(4) High ab.	(5) Low ab.	(6) High ab.	(7) Low ab.
Treatment (intermediate exam)	0.184***	0.343***	0.050	0.360***	0.058	0.215***	0.124**
	(0.057)	(0.083)	(0.053)	(0.078)	(0.055)	(0.081)	(0.053)
Treatment × Credits (demeaned)	0.021*						
	(0.011)						
Macroeconomics	−0.196***	−0.168**	−0.147***	−0.314***	−0.085	−0.177*	−0.067
	(0.057)	(0.084)	(0.050)	(0.087)	(0.056)	(0.099)	(0.071)
Female	−0.054	0.079	−0.142**	0.048	−0.087	0.034	−0.108**
	(0.058)	(0.092)	(0.058)	(0.094)	(0.060)	(0.086)	(0.052)
High School Grade	0.013***	0.019***	0.005	0.011	0.003	0.015**	0.004
	(0.004)	(0.006)	(0.004)	(0.012)	(0.006)	(0.006)	(0.003)
Credits	0.023***	0.043***	0.012	0.042***	0.024***	0.038***	0.010
	(0.007)	(0.013)	(0.008)	(0.009)	(0.005)	(0.009)	(0.006)
Technical Schools	0.024	−0.007	0.058	−0.204	0.191	0.036	0.029
	(0.093)	(0.147)	(0.092)	(0.134)	(0.142)	(0.157)	(0.068)
Lyceum	0.131	0.164	0.030	−0.139	0.274**	0.190	−0.048
	(0.096)	(0.151)	(0.093)	(0.136)	(0.122)	(0.160)	(0.066)
Late Enrollment	0.004	−0.029	0.010	−0.257**	0.197*	−0.080	0.071
	(0.085)	(0.131)	(0.085)	(0.110)	(0.107)	(0.123)	(0.094)
Resident near University	0.050	0.016	0.035	0.037	0.081	0.046	0.025
	(0.056)	(0.095)	(0.053)	(0.092)	(0.056)	(0.087)	(0.056)
Observations	344	166	178	169	175	172	172
Pseudo R-squared	0.255	0.226	0.154	0.267	0.223	0.148	0.172
Log-likelihood	−165.710	−88.771	−70.899	−85.701	−72.155	−101.055	−61.918

Notes: The table reports marginal effects. Robust standard errors are reported in parentheses.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

F_{ij}, which is a dummy equal to 1 for students receiving feedback (that is, treated students undertaking the second module), workload division or commitment, W_i, which is a dummy taking value of one for students undertaking the exam in two separated modules (treated students), a dummy for the second module, D_2, measuring the relative difficulty of this module, and a vector X_i of individual characteristics and a dummy for Macroeconomics. We expect that the workload/commitment has a positive effect on student performance, $\alpha_W > 0$, while according to the literature the feedback effect α_F could be either positive or negative.

In Table 8 we report, on the basis of Eq. (3), the expected performance of treated and control students respectively at modules 1 and 2.

Table 7
IV estimates effective treatment. Heterogeneous effects according to student abilities.

	(1) High ab.	(2) Low ab.	(3) High ab.	(4) Low ab.	(5) High ab.	(6) Low ab.
Treatment (intermediate exam)	0.494***	0.475*	0.513***	0.391	0.393***	0.591***
	(0.074)	(0.256)	(0.058)	(0.265)	(0.112)	(0.115)
Macroeconomics	−0.061	−0.230***	−0.169**	−0.130**	−0.082	−0.198**
	(0.077)	(0.065)	(0.081)	(0.064)	(0.099)	(0.085)
Female	0.011	−0.035	−0.019	−0.070	−0.017	0.017
	(0.077)	(0.115)	(0.078)	(0.065)	(0.080)	(0.083)
High School Grade	0.010*	0.001	0.004	0.002	0.008	0.000
	(0.005)	(0.005)	(0.010)	(0.007)	(0.006)	(0.004)
Credits	0.034***	0.025**	0.033***	0.027***	0.032***	0.025***
	(0.011)	(0.010)	(0.007)	(0.006)	(0.009)	(0.007)
Technical Schools	−0.071	−0.041	−0.125	0.092	0.020	−0.126
	(0.118)	(0.131)	(0.111)	(0.165)	(0.150)	(0.098)
Lyceum	0.056	0.010	−0.027	0.204	0.157	−0.085
	(0.120)	(0.105)	(0.114)	(0.161)	(0.150)	(0.089)
Late Enrollment	−0.020	0.131	−0.061	0.183	−0.081	0.177*
	(0.107)	(0.110)	(0.121)	(0.112)	(0.108)	(0.102)
Resident near University	−0.062	0.054	−0.053	0.073	0.015	−0.006
	(0.086)	(0.063)	(0.077)	(0.065)	(0.081)	(0.068)
Observations	166	178	169	175	172	172

Notes: The table reports marginal effects. Robust standard errors are reported in parentheses.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

Table 8
Student's expected performance by treatment status and module.

Treated, I module:	$E(Y_{T1}) = \alpha_0 + \alpha_W + \phi\bar{X}_i$	since $F_{ij} = 0$, $W_i = 1$; $D_2 = 0$
Treated, II module:	$E(Y_{T2}) = \alpha_0 + \alpha_F + \alpha_W + \alpha_D + \phi\bar{X}_i$	since $F_{ij} = 1$, $W_i = 1$; $D_2 = 1$
Control, I module:	$E(Y_{C1}) = \alpha_0 + \phi\bar{X}_i$	since $F_{ij} = 0$, $W_i = 0$; $D_2 = 0$
Control, II module:	$E(Y_{C2}) = \alpha_0 + \alpha_D + \phi\bar{X}_i$	since $F_{ij} = 0$, $W_i = 0$; $D_2 = 1$

For example, the expected performance for treated students at module 2 is equal to $E(Y_{T2}) = \alpha_0 + \alpha_F + \alpha_W + \alpha_D + \phi\bar{X}_i$ as these students have received feedback, have a reduced workload and are undertaking the second module.

To disentangle the feedback and the workload/commitment effect we consider the following equations:

$$\text{Feedback effect}: \quad [E(Y_{T2}) - E(Y_{T1})] - [E(Y_{C2}) - E(Y_{C1})] = \alpha_F$$

$$\text{Workload division/commitment effect}: \quad E(Y_{T1}) - E(Y_{C1}) = \alpha_W$$

where we exploit the fact that the characteristics (X_i) of treated and control students are on average the same thanks to the random assignment to the two groups.

We base our analysis on a sub-sample of 166 students, excluding from the sample: students who did not sit for the examinations (108 students: 22 treated and 86 control); non complier students, that is, treated students who have shifted to the whole examination (43); students who have taken the intermediate examination but have not taken the second part (27). We exclude these students since for them it is not possible to clearly disentangle the feedback and the workload effects following the framework presented above.

We investigate the effects produced by feedback and workload/commitment both on the grades obtained by students and on their probability of passing the examinations. Since we are using two observations for each student, standard errors have been clustered at student level.

For the sample of students considered in this analysis we do not have censored observations and, as a consequence, we use an OLS model to analyze the effects of interest on *Grade*. Estimates are reported in columns (1) and (2) of Table 9: in the first we do not control for individuals characteristics, which are instead included in the second specification. It emerges that the "workload/commitment" effects are positive and highly statistically significant. Students in the treated group obtain a grade higher of about 4 points (about 0.5 standard deviations of *Grade*). This effect can be interpreted both in relation to the splitting of class workload in two parts – which may help students at better organizing their studying activities – and in relation to the "no procrastination commitment" obtained thanks to the fact that students in the treatment group face more frequent deadlines that may induce them to not procrastinate effort.

On the other hand, the estimates show that the feedback effect is far from being statistically significant.

Similar results are obtained also when we consider as dependent variable the dummy *Pass* using a Probit model. As shown in columns (3) and (4), respectively with and without individual controls, workload/commitment increases the probability of passing the exam of 23.1 percentage points (significant at the 1% level) while the feedback effect is null. The control variables show signs and coefficients similar to those discussed in the previous sections.

Thus, while more frequent examinations turn out to be positive for student performance, this is not due to the fact that students receive a feedback on their performance. The latter result is in contrast with the findings of Bandiera et al. (2009) and Azmat and Iriberri (2010), who instead find a beneficial effect of feedback on student performance. We think that these different results can be explained considering that different types of students may be affected differently from feedback effects. Students considered by Bandiera et al. (2009) are enrolled to a leading UK University that selects among the most able students, while students considered in our experiment are enrolled to a medium level Italian University. If feedback effects are relevant especially for high ability students (as shown by Bandiera et al.), it could be that students included in our sample are not endowed with a sufficiently high level of ability. On the other hand, Azmat and Iriberri (2010) focus on high school students for whom the positive impact of relative performance feedback may be related to social comparison effects, which are typically more relevant in small contexts where students interact repeatedly, such as schools, but less important in tertiary education where courses are taught in large classes and students often know little of each other.

The feedback and workload division effects estimated in Table 9 can be related to the whole treatment effect estimated in the previous sections using as a measure of student performance the average grade of module 1 and 2 ($\bar{Y}_i = (Y_{i1} + Y_{i2})/2$). In fact, using Eq. (3) it is straightforward to show that the whole treatment effect is equal to:

$$E(\bar{Y}_T) - E(\bar{Y}_C) = \frac{(\alpha_0 + \alpha_L) + (\alpha_0 + \alpha_F + \alpha_W + \alpha_D)}{2} - \frac{(\alpha_0) + (\alpha_0 + \alpha_D)}{2} = \alpha_W + \frac{\alpha_F}{2}$$

Estimating by OLS the whole treatment effect (1 observation for each student) on the sub-sample of students considered to disentangle the two effects of interest (not reported), we obtain a coefficient of 4.14 (t-stat = 4.43), which corresponds to the coefficient on workload division (4.367) plus the coefficient (divided by 2) on feedback (−0.452/2).

In Table 10 we report estimation results separately for high and low ability students defined on the basis of their *High School Grade*. We include the full set of controls considered in specifications 2 and 4 of Table 9, but to save space we

Table 9
Feedback and workload division effects.

	(1) Grade (OLS)	(2) Grade (OLS)	(3) Pass (Probit)	(4) Pass (Probit)
Feedback	−0.452	−0.452	−0.003	0.006
	(0.900)	(0.910)	(0.074)	(0.085)
Workload Division	4.058***	4.367***	0.196***	0.231***
	(1.152)	(0.957)	(0.074)	(0.078)
Macroeconomics	0.638	−1.928*	0.058	−0.069
	(1.118)	(1.124)	(0.069)	(0.086)
Module 2	−1.523**	−1.523**	−0.011	−0.016
	(0.647)	(0.654)	(0.054)	(0.061)
Female		−1.417		−0.110
		(0.990)		(0.069)
High School Grade		0.195***		0.011**
		(0.060)		(0.004)
Credits		0.560***		0.027***
		(0.095)		(0.007)
Technical Schools		1.205		0.011
		(1.708)		(0.115)
Lyceum		2.690*		0.062
		(1.606)		(0.111)
Late Enrollment		−0.229		−0.012
		(1.525)		(0.104)
Resident near University		−0.461		0.011
		(1.066)		(0.073)
Constant	16.588***	−9.981*		
	(1.023)	(5.551)		
Observations	332	332	332	332
R-squared	0.065	0.364		
Pseudo *R*-squared			0.031	0.169
Log-likelihood			−215.447	−184.648

Notes: Observations at module-student level: 2 observations for each student. OLS estimates in columns 1 and 2. In columns 3 and 4 we report marginal effects of a Probit model. Robust standard errors, clustered at student level, are reported in parentheses.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

do not report the coefficients on these variables. In columns 1 and 2 we report results obtained considering as dependent variable *Grade*. We find that the feedback effect is not relevant neither for high ability students nor for low ability ones, while a positive workload effect emerges for both the two groups of students. In the same direction go the results obtained considering the effects on the probability of passing the exams (columns 3 and 4). These findings are robust to the use of alternative measures of ability (*Composite Ability* and *Credits*) to split students in a high and a low ability group (not reported).

7. Are students substituting effort?

In this section we investigate if treated students – thanks to the opportunity to undertake the intermediate examination – have provided more effort and achieved a better performance or whether they have focused on the targeted examinations devoting less effort in studying activities related to other examinations.

In order to evaluate this aspect we analyze student performance on all the examinations students have to pass in the academic year considered. We stack data with the

Table 10
Heterogeneity in the feedback and workload division effects.

	(1) Grade High ab.	(2) Grade Low ab.	(3) Pass High ab.	(4) Pass Low ab.
Feedback	−0.018	−1.159	0.057	−0.090
	(1.072)	(1.723)	(0.097)	(0.136)
Workload division	3.907***	4.031***	0.183**	0.210*
	(1.178)	(1.591)	(0.087)	(0.127)
Observations	210	122	210	122
R-squared	0.364	0.311		
Pseudo *R*-squared			0.170	0.177
Log-likelihood			−104.212	−68.476

Notes: Observations at module-student level: 2 observations for each student. OLS estimates in columns 1 and 2. In columns 3 and 4 we report marginal effects of a Probit model. Robust standard errors, clustered at student level, are reported in parentheses.
* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

Table 11
The probability of passing targeted and non-targeted exams. Linear probability model.

	(1) First year	(2) Second year
Treatment	0.131*	0.260***
	(0.070)	(0.069)
Treatment × Exam 1	−0.110	−0.248***
	(0.074)	(0.075)
Treatment × Exam 2	−0.154	−0.362***
	(0.095)	(0.091)
Treatment × Exam 3	−0.131*	−0.236***
	(0.078)	(0.077)
Treatment × Exam 4	−0.090	−0.256**
	(0.078)	(0.099)
Treatment × Exam 5	−0.161**	−0.258***
	(0.082)	(0.092)
Treatment × Exam 6	−0.136*	−0.367***
	(0.076)	(0.088)
Treatment × Exam 7	−0.117	−0.239***
	(0.080)	(0.080)
Exam 1	0.433***	−0.147***
	(0.051)	(0.046)
Exam 2	−0.082	0.507***
	(0.067)	(0.061)
Exam 3	0.443***	−0.053
	(0.057)	(0.046)
Exam 4	0.412***	0.493***
	(0.058)	(0.072)
Exam 5	0.247***	0.627***
	(0.061)	(0.060)
Exam 6	−0.155***	0.427***
	(0.056)	(0.069)
Exam 7	−0.227***	0.133**
	(0.056)	(0.055)
Implied Treatment Exam 1	0.021	0.012
	(0.040)	(0.031)
Implied Treatment Exam 2	−0.023	−0.101
	(0.070)	(0.078)
Implied Treatment Exam 3	0.000	0.024
	(0.041)	(0.052)
Implied Treatment Exam 4	0.041	0.004
	(0.041)	(0.074)
Implied Treatment Exam 5	−0.030	0.002
	(0.063)	(0.064)
Implied Treatment Exam 6	−0.005	−0.107
	(0.065)	(0.078)
Implied Treatment Exam 7	0.014	0.022
	(0.061)	(0.074)
Observations	1544	1208
R-squared	0.344	0.326

Notes: The dependent variable is the dummy *Pass*. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the student level. In all the regressions we control for Female, High School Grade, Type of High School attended, Late Enrollment, Residence near the University.

* Coefficients are statistically significant, at 10% level.
** Coefficients are statistically significant, at 5% level.
*** Coefficients are statistically significant, at 1% level.

aim to use student-class level observations and deal separately with students of the first year (those attending Microeconomics) and students of the second year (students attending Macroeconomics).

In the first year students have to pass 8 examinations[9]: we have for these students 1544 observations (193 students × 8 classes); students have to pass 8 examinations also in their second year[10]: we end up with 1208 observations (151 students × 8 classes). Our dependent variable, *Pass*, is a dummy equal to 1 when the student passed a given exam, and 0 otherwise.

We use a Linear Probability Model to estimate the following equation:

$$Pass_i = \beta_0 + \beta_1(Treatment_i) + \phi X_i + \psi_1 E_{1i} + \cdots + \psi_7 E_{7i} + \pi_1(Treatment_i \times E_{1i}) + \cdots + \pi_7(Treatment_i \times E_{7i}) + \varepsilon_i \quad (4)$$

where the probability of passing each examination is related to individual characteristics *X*, a dummy E_j (j = 1, . . ., 7) for each examination (leaving as omitted category the eighth examination – Microeconomics or Macroeconomics according to the cohort of students considered) to control for unobserved factors such as the difficulty of the subject or the instructor's grading standard. The dummy *Treatment* allows us to identify the effect of treatment on the target examination (β_1) while the coefficients π_j on the interaction terms ($Treatment \times E_j$) inform us on whether treated students obtain a worse performance compared to control students in non targeted examinations. In fact, the performance of treated students in examination *j* is given by $\beta_1 + \pi_j$.

Estimates are reported in Table 11. In column 1 we report results considering first year students attending the Microeconomics class, while in column 2 are presented the estimates for second year students attending the Macroeconomic class. The coefficients on *Treatment* are, as expected, in line with those shown in Table 3 (columns 3 and 4). On the other hand, we do not observe any significant difference among students in treated and control groups in the performance on non-target examinations. As shown at the foot of the table, where are reported the linear combinations of $\beta_1 + \pi_j$ and the respective standard errors, the coefficients are never statistically significant.

Therefore, our estimates show that the improvement in student performance at the targeted examination is to be related to a higher student effort and is not driven by substitution effects.

8. Concluding remarks

Policymakers and researches often debate about students' performance and on policies that may help at boosting them. The improvement of students' performance is important both when considering primary and secondary education and when looking at tertiary education since the skills acquired by undergraduate students are crucial for their success in the labor market. In addition, since in many countries, and particularly in Italy, tertiary education is characterized by high drop out rates and excessive duration of the academic career, the improvement of students performance may translate in a reduction of the number of students dropping out from University studies.

[9] Business Administration 1, Public Law, French 1, Computer Sciences, English 1, Mathematics, Statistics, Microeconomics.

[10] Trade Law, Business Administration 2, Private Law, French 2, English 2, Financial Mathematics, Accounting, Macroeconomics.

Whereas a widely investigated strategy to improve student performance is that based on educational resources, a considerable impact may derive also from changes in the organization of teaching activities and in evaluation practices. Recently, a number of papers have shown that providing students with feedback information both on their interim performance that on the performance of their peers helps at increasing students' achievements (Azmat & Iriberri, 2010; Bandiera et al., 2009). In this vein, the aim of our paper has been to investigate the effect of more frequent examinations allowing students to obtain a number of beneficial effects, deriving not only from the provision of feedback but also from the division of the class workload from the commitment allowed by recurrent deadlines.

To analyze these effects we have carried out a randomized experiment involving undergraduate students enrolled at an Italian University and attending two introductory economics classes. Students included in the treated group were allowed to undertake an intermediate exam and were informed about the results obtained, while students in the control group were allowed to undertake exclusively the final examination at the end of the classes (as established by the University rules).

From our analysis it emerges that students undertaking the intermediate examination obtain a better performance both in terms of probability of passing the exam and of grades obtained. Treated students have a probability of 20 percentage points higher of passing the exam and their grades are 4–5 points higher. These positive effects are mainly concentrated among students endowed with higher abilities. Moreover, we find that the better performance obtained by treated students at the targeted examination seems not to be driven by substitution effects.

The design of our experiment allowed us to disentangle the effect deriving from more frequent examinations in a "feedback provision" effect and a "workload division or commitment" effect. Our estimates show that the positive impact of the intermediate examination is entirely due to the workload division or commitment effect, while it turns out that the feedback provision has no positive effect on performance.

According to our results the recent law that in Italy has reorganized classes in longer modules defining a maximum number of examinations may produce negative effects, increasing the duration of students' academic career.

Acknowledgements

We would like to thank the University of Calabria for providing access to the administrative data and the Dean of the Faculty of Economics for allowing us to carry out the experiment involving students enrolled at the Business and Administration Degree. We also thank Sara Laurita and Antonella Caparelli who assisted us with the use of the data, Marco De Benedetto, Filippo Domma, Francesca Gioia, Sabrina Giordano, Pierfrancesco Perri, Michela Ponzo, two anonymous referees and seminar participants to the Second International Workshop on Applied Economics of Education, Catanzaro, 2011, for helpful comments.

References

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

Angrist, J., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, Deadlines, and Performance: Self-Control by Precommitment. *Psychological Science*, *13*(3), 219–224.

Aoyagi, M. (2010). Information feedback in a dynamic tournament. *Games and Economic Behavior*, *70*(2), 242–260.

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, *94*(7–8), 435–452.

Bandiera, O., Larcinese, V., & Rasul, I. (2009). *"Blissful Ignorance? Evidence from a natural experiment on the effect of individual feedback on performance", Policy Research Working Paper Series, IZA*.

Bloom. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, *8*, 225–246.

Burger, N., Charness, G., & Lynham, J. (2011). Field and online experiments on self-control. *Journal of Economic Behavior & Organization*, *77*(3), 393–404.

Deck, W. (1998). *The effects of frequency of testing on college students in a principles of marketing course*. PhD Dissertation, Digital Library and Archives.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, *4*, 3895–3962.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*.

Eriksson, T., Poulsen, A., & Villeval, M. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, *16*, 679–688.

Freeman, R., & Gelber, A. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, *2*(1), 149–164 (16).

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Lizzeri, A., Meyer, M., & Persico, N. (2002). *The Incentive Effects of Interim Performance Evaluations* USC FBE APPLIED ECONOMICS WORKSHOP.

Perry, M., & Reny, P. J. (1999). On the failure of the linkage principle in multi-unit auctions. *Econometrica*, *67*, 895–900.

Schacter, J., & Thum, Y. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, *23*(4), 411–430.

Schwerdt, G., & Wuppermann, A. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review*, *30*(2), 365–379.

Tuckman, B. (1997). Using tests as an incentive to motivate procrastinators to study. *Journal of Experimental Education*, *66*, 141–147.

Yildirim, H. (2005). Contests with multiple rounds. *Games and Economic Behavior*, *51*(1), 213–227.