Warwick Economics Research Paper Series

Knowing who you are: The Effect of Feedback Information on Short and Long Term Outcomes

Sofoklis Goulas and Rigissa Megalokonomou

October, 2015
Series Number: 1075
ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Knowing who you are: The Effect of Feedback Information on Short and Long Term Outcomes

Job Market Paper

Sofoklis Goulas
Department of Economics
University of North Carolina at Chapel Hill
goulas@unc.edu

Rigissa Megalokonomou
Department of Economics
University of Warwick
r.megalokonomou@warwick.ac.uk

Abstract

We study the effect of disclosing relative performance information (feedback) on students' performance in high-school and on subsequent university enrolment. We exploit a large scale natural experiment where students in some cohorts are provided with their national and school relative performance. Using unique primary collected data, we find an asymmetric response to the relative performance information: high achieving students improve their last-year performance by 0.15 standard deviations whereas the last-year performance of low achieving students drops by 0.3 standard deviations. The results are more pronounced for females indicating greater sensitivity to feedback. We also document the long term effect of feedback provision: high achieving students reduce their repetition rate of the national exams, enrol into 0.15 standard deviations more popular University Departments and their expected annual earnings increase by 0.17 standard deviations. Results are opposite for low achieving students. We find suggestive evidence that feedback encourages more students from low-income neighborhoods to enrol in university and to study in higher-quality programs indicating a potential decrease in income inequality.

Keywords: feedback, relative performance, university admission, rank, gender differences, income inequality
JEL Classification: I23, J21

1 Introduction

Improving pupils' attainments has been an important issue for policy makers and academics alike. In an effort to improve students' grades, education policies have focused on improving school inputs such as reducing class size (Angrist and Lavy 1999, Krueger 1999), improving the quality of teachers (Chetty et al. 2014, Rothstein 2010, Aaronson et al. 2007), extending the term length (Card and Krueger 1992) and improving the quality of the peer group a student is exposed to (Lavy et al. 2012, Zimmerman 2003, Hoxby 2000a). However, little is known about whether providing social comparison information improves students' performance. Manipulating the availability of social comparison information could be significantly less costly than the above mentioned interventions.

This paper presents a theoretical motivation and empirical evidence of whether providing high school students with social comparison information regarding their performance in externally marked high stake exams affects future performance in similar exams. Our analysis relies on the fact that different cohorts have different policies regarding the provision of social comparison information.

We exploit a large scale natural experiment that took place in Greece in 2005. The policies we observe differ based on whether students receive information about their ordinal rank position at the end of the eleventh grade. Until 2005, all students had to take national exams in two adjacent grades; one year before graduation from high-school and the year they graduated from high school (feedback regime). In this regime, each student's performance in the eleventh grade exams was publicly announced, giving students the opportunity to calculate their national and school rank. Thus, we define feedback as the information of one's performance in comparison to their peers in school and nationwide. In the feedback regime, students could compare themselves to others allowing for social comparison. Disclosing information about someone's ordinal rank could change people's behaviour when they make comparisons with others (Card et al. 2012). When students knew their performance in the eleventh year exams, they could translate their hours of effort into exam result. Knowing their relative performance could also affect the amount of effort students decide to exert towards their twelfth grade performance (Ertac 2005). Students' performance in the twelve grade (final year) national exams is the most important determinant for University admission in this setting.

After 2005, the penultimate year (eleventh grade) national exams are abolished and replaced by school exams. This means that after 2005, penultimate year students sit exams on the same subjects as before but they now receive report cards with their own grades only. As a consequence, they no longer receive information about their penultimate year *relative* performance. These cohorts -as the previous ones- sit national exams in the twelfth grade that will determine their post-secondary placement. However, they have been imposed a loss of feedback information regarding previous performance in similar exam (non-feedback regime).

Using new data on school performance, school quality and national exams for university admission, we test the hypothesis that students' final year exam (twelfth grade) performance is independent of the feedback regime. Conditional on their tenth grade performance, we compare the final year performance of students across feedback regimes. After controlling for students' characteristics, we identify the effect of feedback provision on their short term (academic performance in the University entrance exams) and their long term outcomes (repetition of national exams one year after graduation, popularity of University Department admitted to and expected annual earnings).

Our first finding is that high achieving students perform better in externally graded exams when they are aware of their relative performance in the school and nationwide. Feedback information on past performance improves the next period's exam performance of the better students by 0.2 standard deviations and their relative national rank by 4-6 percentiles. This is of comparable magnitude to being taught by a teacher 1.5-2 standard deviations above the average (Chetty et al. 2014, Hanushek et al. 2005) or to reducing the class size by 15 percent. (Angrist and Lavy 1999, Krueger 1999). Additionally, we find evidence that the performance of students in the lower percentiles deteriorates when feedback is provided. In particular, their consecutive year performance declines by 0.3 standard deviations and their national rank decreases by 6-8 percentiles. To build intuition here, we consider that knowing how someone performs relatively to others in the same task affects his motivation to exert less or further effort in the next time period. (Ertac 2005)

Our second finding reports the response of males and females to feedback at different parts of the ability distribution. High achieving students of both genders respond positively to positive feedback and low achieving students of both genders respond negatively to negative feedback. However, females seem to be considerably more sensitive to feedback at all parts of the ability distribution than males. Our results are consistent with the existing literature findings regarding gender differential response to feedback information due to initial different levels of self-confidence (McCarty 1986).

Our third finding is that the provision of feedback changes the matching of students to University Departments. First, we rank all University Departments (program) based on popularity and we construct a program popularity list reporting the most popular programs (like engineering and medicine) and moving to the least popular ones (like geo-technology and environmental studies). We find that feedback provision makes high achieving students move up the program popularity ladder by 30 positions which is a 0.15 of a standard deviation. On the other hand, low achieving students move down the program popularity ladder by 35 positions which is a 0.18 of a standard deviation, when feedback is provided. Using the Labor Force Survey information, we find the annual earnings of older people who work in each occupation and we map them to University Departments. When the social comparison information is disclosed, we find that high achieving students experience an increase in expected earnings by 0.13 standard deviations. Further, feedback provision imposes a decrease in expected earnings by 0.23 standard deviations for low achieving students.

Additionally, we find suggestive evidence that feedback alters the socio-economic background composition of students who manage to get admitted to the top programs. More students from low income neighborhoods get admitted to the most prestigious programs with the highest expected earnings after graduation (like engineering and law), when feedback information is provided. This implies that feedback information encourages social elevation motivated by students from low income families.

We believe that this paper has two main contributions. First, this is the first large scale study that documents the long term effects of providing relative performance information in an education setting. In particular, we document the direction and size of the effect of feedback provision on long term outcomes such as repetition rates of the national exams, students' post-secondary placement and expected earnings. We contribute to the literature by providing evidence that knowing someone's rank position within his senior high school or throughout the country might have long lasting effects and change students' career path. Thus, rank information can be considered a new factor in the education production function [1]. In a concept where parents can choose

[1]Other determinants of the educational production function that have been studied are: studies on class size (Angrist and Lavy 1999, Krueger 1999, Hoxby 2000b), teachers' training and certification (Angrist and Lavy 2001, Kane et al. 2008), quality of teacher (Rockoff 2004, Rivkin et al. 2005), tracking (Duflo et al. 2011), peer

the best school for their children, that could have important policy implications. Our findings imply that when the relative performance information is released, being in a school with higher-achieving peers might not always be optimal for students. In particular, students might be benefited by going to schools where they are among the high-performing students i.e. schools with lower than them performing pupils. Imagine two students of the same high ability. The one is in school X and gets a very high rank. The other one is in school Y with higher-achieving students and gets a lower rank. The student who was among the best in his school X will do better in the subsequent University entrance exams and enrol into a better program than the student who is in school Y surrounded by high achieving students.

Secondly, we provide evidence that a low cost instrument -like providing rank information- has the potential to affect students' education achievements. Our estimates are at the lower end of those compared to the literature on improving school inputs. Nonetheless, all the interventions studied so far (improving teachers quality, reducing class size, enhancing the peer quality group) are significantly more costly than manipulating the availability of social comparison information. However, disclosing information about ranks is rare as it is not a standard practice for teachers or principals to discuss rankings. Thus, the information treatment that we study in this paper is unique. We exploit a very special setting where high school students receive explicit information about their relative position in two reference groups; school and country. Although someone may generally observe his own perspicacity, they do not generally observe everyone's performance in the school and the country in order to deduce their rank position. In our setting, the social comparison group widens when students receive information about the peers' performance in the school and nationally. Thus, we are able to separately identify the effect of knowing someone's rank position in each of these two groups.

We also discuss the two most prevailing mechanisms that could explain how students react to the social comparison information and are related to students: 1) learning about own relative ability and/or 2) learning about the quality of the school. The mechanism that best accommodates all our findings is the first one with the second helping us rule out alternative interpretations of the results. The first mechanism is related to the importance of non-cognitive skills on educational outcomes and especially self-

effects (Hoxby 2000a, Lavy et al. 2012), non-cognitive skills (Heckman et al. 2006), classroom instructional time (Lavy 2015).

perception. The importance of non-cognitive skills is well established in the literature (Brunello and Schlotter 2011, Heckman et al. 2006, Kautz et al. 2014)

In the recent years there has been an increasing interest in the economic literature of feedback information provision on exam performance.[2] Bandiera et al. 2008 examine the effect of feedback information on students' future absolute performance using data for University students registered to Departments with different feedback policies. In that study, feedback is defined as the knowledge of someone's absolute performance in the midterm exam in period one and before students exert effort on their essay in period two. The authors find that the effect of feedback is positive for all students and more pronounced for more able students. However, their study refers to feedback involving own performance. The provision of feedback regarding relative performance has not received much attention.

The paper which is most closely related to ours is a study by Azmat and Iriberri 2010. The authors examine the effect of relative performance feedback on students' future absolute performance. They exploit a natural experiment that took place in a high school, where for one year only students received information about the average class score in addition to their own performance. Their findings suggest that feedback improves the performance of all students in the subsequent test. They do not find differential effects by gender along the ability distribution. A key difference to our work is that they use a small sample of one high school while we use a sample of 134 senior high schools nationally representative in many dimensions. Another important difference is that Azmat and Iriberri 2010 investigate the effect of providing information about someone's relative position within the class only. We contribute to the literature by examining the effects of providing broader social comparison information about someone's school and national rank. More recently, Murphy and Weinhardt 2014 examine the effect of knowing one's ordinal rank position in exam results on future exam performance. They find large and robust effects of being highly ranked in primary school on secondary school achievement. Their study also reports that boys are more affected by knowing their ordinal rank than girls. In their setting, students

[2]The relative feedback information has been studied in the tournament literature. Some studies find that relative performance information has a positive effect for all participants in tournaments and piece rate payment schemes (Hannan et al. 2008). On the other hand, some other studies find mixing results. Barankay 2012 uses data on furniture salesmen's effort and finds that feedback has negative effects on the low performing employees.)

figure out their rank within their class from social interaction with their classmates. In our setting, the information set is broader and is provided by the principal. Students receive explicit information regarding their rank position within the school and nationwide which facilitates the policy recommendations.

An interesting question is if the effects are driven by students, parents or teachers. It is almost impossible to disentangle if the effect is coming from the students or the parents, but we can rule out the possibility that there is a sorting into schools by parents due to the structure of the system. We are also able to rule out the possibility that teachers are driving the results because the national exams are externally marked and teachers have no way to affect these grades. Additionally, teachers cannot allocate students into classes in a way that facilitates sorting because it is prohibited by the law.

The paper is organized as follows. Section 2 presents a theoretical model for the individual's behaviour and motivates the empirical investigation. Section 3 provides a brief description of the institutional setting and the data. Section 4 sets out our empirical strategy. Section 5 presents the main results on short and long term outcomes and discusses heterogeneous feedback effects by ability, gender, track and neighborhood income. Section 6 discusses the threats to identification and reports further robustness checks. Finally, in Section 7 we conclude and discuss possible policy implications.

2 Theoretical Framework

We adapt a theoretical model proposed by Ertac 2005 [3] where students have imperfect information about their own ability.

In the non-feedback regime eleventh graders sit school exams and they receive information about their own performance only. In the feedback regime, they receive information about their own performance and about the school and cohort average performance. Students take exams in two time periods; the eleventh and the twelfth grade. Students' performance in the eleventh grade depends on their ability and the easiness of the task. This performance provides them with some information about ability and

[3] Ertac 2005 presents a principal-multiple agents model where agents have imperfect information about their abilities under multiple types of contracts. The model is also used by Azmat and Iriberri 2010. The natural experiment they study gives students information about the average grade of the class, while here the social comparison information refers to the average school and cohort grade.

easiness of the exam [4]; we will refer to that as the private signal s_i. The ability of a student is denoted by $\alpha_i > 0$ and α_i's are independent draws from the same distributions and independent of task difficulty. All distributions are common knowledge.

When signals coming from the eleventh grade are realized, students update their beliefs about their ability and decide their subsequent effort. The amount of effort students decide to exert in the twelfth grade determines their final year's scores. Period 1: This is the learning stage. Students receive a noisy signal about their ability:

$$s_i = \alpha_i + c \quad , i = 1, 2, ...$$

$$\alpha_i \sim N(\bar{\alpha}, \sigma^2), \bar{\alpha} > 0$$

$$c \sim N(0, \psi^2)$$

$$cov(\alpha_i, c) = 0, \quad cov(\alpha_1, \alpha_2 = 0)$$

This signal (s_i) depends on student's i ability level (α_i) and a shock that is common to all students [5]ie. the easiness of the exam (c). We also assume that α_i and c are normally distributed and α_i and s_i are jointly normally distributed. In the feedback regime, students in each school or nationwide also observe the average signal:

$$\bar{s} = \frac{\sum_{i=1}^{N} s_i}{N} = \frac{\sum_{i=1}^{N} (\alpha_i + c)}{N} = \frac{\sum_{i=1}^{N} (\alpha_i)}{N} + c$$

The type of the signal each student receives, affects student's perceived belief about his own ability in the first period and his belief about his own ability determines the amount of effort he chooses to exert in the second period. Then we find student's i expectation of his own ability conditional on the type of signal he receives [6]. In the non-feedback regime the student observes his own performance in the school exams. His expected ability given the observed signal is:

$$E(\alpha_i | s_i) = \bar{\alpha} + \frac{\sigma^2 (s_i - \bar{\alpha})}{\sigma^2 + \psi^2}$$

[4]In the feedback regime c corresponds to the easiness of the national exams. In the non-feedback regime c corresponds to the easiness of the school exam.

[5]In the school or the cohort depending on the feedback or non-feedback regime.

[6]Using the properties of the multivariate normal distribution, we find that

$$\begin{pmatrix} \alpha_i \\ s_i \\ \bar{s} \end{pmatrix} \sim \begin{pmatrix} \bar{\alpha_i} \\ \bar{\alpha_i} \\ \bar{\alpha_i} \end{pmatrix} \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2/N \\ \sigma^2 & \sigma^2 + \psi^2 & (\sigma^2 + N\psi^2)/N \\ \sigma^2/N & (\sigma^2 + N\psi^2)/N & (\sigma^2 + N\psi^2)/N \end{pmatrix}$$

In the feedback regime, the student receives also the social comparison information :

$$E(\alpha_i|s_i, \bar{s}) = \bar{\alpha} + \frac{(\sigma^2 + \psi^2)(s_i - \bar{\alpha}) - \psi^2(\bar{s} - \bar{\alpha})}{\sigma^2 + 2\psi^2}$$

The higher the private signal a student receives, the higher is his belief about perceived ability. If the average signal is observed, then the belief about ability decreases with it.

Period 2: Following the realisation of the signals, in the second period students choose the effort to exert (e_i). Students' objective is to maximize the second period performance (q_i) after choosing the effort to exert. Assuming that the performance production is a linear function in effort [7] and that effort and ability are complements in performance[8] it follows that: $q_i = \ e_i$. There is also a cost associated with the effort exerted that is $c(e_i)$ and is increasing in effort and convex. [9] In the second period, students choose the effort level $e_i > 0$ in order to maximise their last year's utility function. In the absence of the social comparison information students receive only the private signal and they maximise:

$$u^{NF} = E[p_i(\alpha_i, e_i) - c(e_i)|s_i] = E[\alpha_i|s_i]e_i - c(e_i)$$

and the F.O.C simplifies to $E[\alpha_i|s_i] - c'(e_i^{NF*}) = 0$ (1)

In the feedback regime where social comparison information is provided the student observes the average signal and maximises:

$$u^F = E[p_i(\alpha_i, e_i) - c(e_i)|s_i, \bar{s}] = E[\alpha_i|s_i, \bar{s}]e_i - c(e_i)$$

and the F.O.C simplifies to $E[\alpha_i|s_i, \bar{s}] - c'(e_i^{F*}) = 0$ (2)

Given that the F.O.Cs are sufficient, we will compare the optimal effort levels in the two regimes. The conditional expectation of ability is independent of effort while the second term in (1) and (2) is an increasing function of effort. That means that an increase in the beliefs about ability -a higher self confidence level- leads to an increase in the optimal effort level. The comparison of the F.O.Cs for the two regimes simplifies to the comparison of the conditional expected abilities.

$$E[\alpha_i|s_i, \bar{s}] = E[\alpha_i|s_i] \quad if \quad s^* = (\bar{s} - \bar{\alpha})\frac{N(\sigma^2 + \psi^2)}{\sigma^2 + N\psi^2} + \bar{\alpha}$$

[7] The predictions of the model do not change if the performance function is not linear in effort.

[8] $\frac{dqi}{d\alpha_i de_i} > 0$

[9] $(c'(e_i) > 0, c''(e_i) > 0, c'(0) = c''(0) = 0)$

Thus, if $s_i > s^*$ then $e^{F*} > e^{NF*}$ and if $s_i < s^*$ then $e^{F*} < e^{NF*}$. [10]. Students with signal above (below) s^* will put in more (less) effort, when feedback is provided. If $\bar{s} = \bar{\alpha}$ then the exam is neither hard nor easy. If $s^* = \bar{\alpha}$ which means that $s^* = \bar{s}$ and the average signal equals the average ability level and $e^{F*} = e^{NF*}$. However, if $s^* > \bar{\alpha}$ then $s^* > \bar{s}$ and if $s^* < \bar{\alpha}$ then $s^* < \bar{s}$. That means that if the signal is above the average signal then students will exert more effort when feedback is provided. Similarly, if the signal is below the average signal then students will exert less effort when feedback is provided.

If $\bar{s} > \bar{\alpha}$ then the exam was hard and the signal needed in order for students to exert more effort is higher than the average signal($s^* > \bar{s}$). If $\bar{s} < \bar{\alpha}$ then the exam was easy and the signal needed in order for students to exert more effort is lower than the average signal($s^* < \bar{s}$)

Let us summarize now the main hypothesis about the effect of the eleventh grade social comparison information on the twelfth grade performance.

Null Hypothesis: Students do not react to the social comparison information

That would suggest that students are not uncertain about their ability or that students have already figured out their relative performance information and the explicit addition of it is redundant or that the private signal that students get in the feedback regime equals the average signal.

Alternative Hypothesis: Positive effect on performance for high ability students and negative effect on performance for low ability students

That would suggest that students will react differently to feedback. Based on the model, high ability students will perform better when the social comparison information is provided because they have been encouraged by their period one performance. On the other hand, low ability students will perform worse when the social comparison information is provided because they have been discouraged by their eleventh grade performance.

Notice here that there is no pass-fail scheme and students do not try to achieve a performance threshold. University cutoffs are determined endogenously based on demand and pre-specified supply of seats. In other words, the model makes these predictions based on the fact that ability and effort are complements in the production

[10] $\frac{N(\sigma^2+\psi^2)}{\sigma^2+N\psi^2} > 1$ provided than $N >= 2$

function. [11]

3 Institutional Setting and Data

3.1 Institutional Setting

In Greece, all students in secondary education are obliged to take the national exams to have access to tertiary education. Students sit these national exams in specific subjects on specific dates every year and the questions asked are the same for all students in that cohort. The national exams are externally marked. All universities are public and the admission procedure is run exclusively by the Ministry of Education. University admission in Greece is based on the "admission grade". The admission grade in both regimes is a weighted average of the grades a student gets in the national exams (70% weight) and the school grades (30% weight). The school grade of every subject is the average of the term grades. Only final year students can participate in the university admission procedure. Admission is made in a specific university department. All students are examined on five general education core modules. On top of that, students are examined on Elective subjects that are determined by the "speciality" or the "track" they choose at the beginning of the twelfth year.

The admission grade of a student in the non-feedback regime depends entirely on students' performance in the twelfth grade. It is a combination of the national exams (70%) and the school exams (30%). In the feedback regime, students' performance in the eleventh grade could take some weight (30%) in the calculation of the admission grade. That is the case if and only if the overall performance of a student in the eleventh grade is better than that in the final year exams [12]. The overall performance of a student in each grade is calculated again as a combination of the national exams (70%) and the school exams (30%). The results of the penultimate year exams could not be used in any other way in the university admission procedure. So students have incentives to perform well in the eleventh grade national exams but that is not enough to secure a specific University placement. Given that the number of University seats is

[11] In a different setting where University cutoffs are pre-determined, effort and ability could be substitutes in the production function. In that case, a student who is above average in the eleventh grade may choose to exert less effort in the twelfth grade in order to achieve a specific performance threshold.

[12] In this case, the overall performance of a student in the twelfth grade takes a weight of 70% and the overall performance of a student in the eleventh grade takes a weight of 30% in the calculation of the admission grade.

fixed, scoring a particular admission score does not guarantee admission to a specific University Department. It highly depends on competition with other students in that cohort.

First, students take the final-years exam. Then, students' admission grades are announced. Then, every student makes and submits to the Ministry of Education a preference list of university departments he would like to be admitted to in that year. If a student is admitted to a University Department in a higher place in his preference list he cannot be admitted to those below that. That makes students to be very careful in constructing their preference list. The only way a student can flee from the university admission procedure is to deny submitting a list of preferences. Every university department admits a pre-specified number of students. Then, each department admits the best students that have included this department in their preference list. All students are compared to each other according to their admission grades and every successful candidate is admitted to the first department in his list where there is an available place and every student with higher admission grade has already been allocated. The rest of the students are denied admission at that year.

At the end, every department announces the grade of the last student it admitted in that year. This grade is considered to be the "bottom grade" or the "cutoff grade" in that year for each university department. More popular departments exhibit higher cutoff grades. Students are aware of the " cutoff grades" of the previous years when they construct their preference list. The ranking of university departments according to their cutoff grades appears to stay largely unchanged, year after year, and this represents the students' valuations for these departments. It's not possible to defer someone's admission. Some students that have not been admitted to the university department they wanted to may decide to retry admission a year (or more) after graduation using their school grades in the admission procedure and retaking national exams in all subjects. Those students usually do not attend any school/college or pursue any job or do military service after graduation and before the next admission period.

3.2 How does feedback work?

Knowing one's own relative performance might affect the amount of effort students exert with regard to a certain objective. In the context of our study, students' objective is the maximization of their score and/or rank at the end of senior high school.

Consider a student in the treated group. This student is of certain ability and her objective is to maximize her score given hers and everyone else's choice of effort. In the world of this experiment, students compete with each other over access to a limited number of places in higher education. At the end of the penultimate year everyone takes standardized exams in some subjects with external interlocutors and at least two anonymous external markers per subject.

Then two mechanisms are in action: First, the scores of all students across the country become public knowledge. In particular, the names of all students who take the national exams together with their national exam results and the cohort's average national score are announced in the newspaper. So each student could calculate her distance from the cohort's average score and derive her relative rank in comparison to her cohort.

Second, everyone's results within the school become public knowledge as the names and detailed grades are printed and pinned on boards at the entrance of every school. Humans are social beings and social comparison is an indispensable part of bonding among adolescents. Once school starts again next year, our student has an idea of how well she can do given a specific level of effort when national exams come around again. Most importantly, she knows how well she can do relative to her schoolmates given hers and their choice of effort.

For the sake of comparison let's consider a student in the control group of our study. Given his ability he chooses an effort level at the penultimate year of senior high school in order to succeed in the end of the year school exams necessary to advance to the twelfth grade. Teachers coordinate to cover the same material and usually give the same exam questions intra-school. Before the summer break in the penultimate year, our student sits exams on the same five subjects and receives a written report from school with his own grades. When he reaches twelfth grade, he has access to the same material, study guides and past exam papers as any student in the treated group. He is only unaware of how his schoolmates and his cohort did relative to him in the penultimate year final exams. Table 3 reports the summary statistics of the variables of interest across the two regimes. Most of the differences seem to be significantly different zero but they are either very small or economically non meaningful. The exact timing is presented in Figure 2.

3.3 Data Collection

The transition from high school to higher education is based on a centralised allocation of students to University Departments. The admission procedure is run exclusively by the Ministry of Education, which collects data on students' performance only if they are relevant to the calculation of the admission grade. In order to study the effect of disclosing rank information, we need a prior measure of performance that is not affected by the provision of the social comparison information i.e students' tenth grade performance. Students' performance in the tenth grade is not centrally collected and can only be found in the school archives. Thus, we visited senior High schools across the country and we have constructed a database of detailed student performance in every subject throughout senior high school. In particular, we use data collected from a large sample of 134 schools across the country. Our novel dataset combines information from various sources:

1. Administrative data obtained by the Ministry of Education regarding the twelfth grade performance of all students who sat the twelfth grade national exams from 2003 to 2009. This dataset contains student level information about gender, national and school exam results in each subject nationally examined in twelfth grade, name of senior High school attended, year of birth and graduation year from senior High school, speciality chosen at the beginning of twelfth grade. It also contains University admission related information such as the University Department each student got admitted to, number of applications made to University Departments and the reported ordinal preference position of the University Department admitted in someone's preference list. The dataset refers to the period 2003-2009 and gives us information about 435.589 students.

2. As the Ministry does not collect information on students' tenth grade performance, we collected this information directly from the schools.[13] More specifically, we have

[13]The tenth grade performance data are recorded in each school's archives either in their computers or in their history books. In most schools the data for all the years were extracted from their computers. There were cases-especially for the data referring to the first years of our sample period- where we photocopied pages from the history books in schools' storage area.

physically visited and collected data from 134[14] public, experimental [15] and private schools both near big cities and in the countryside (this number corresponds to around 10 % of the school population). We exclude the evening schools[16] from our analysis because they differ in many aspects from the other types of schools[17]. This dataset includes information about school and/or national exam results in tenth, eleventh and twelfth grade in all subjects, indicators for gender, a class indicator, graduation year, year of birth, speciality chosen at the beginning of the eleventh and twelfth grade and a unique identification code for each student that stays the same throughout senior high school. We have had short interviews with the principal of every school in our sample to find out about any effects potentially affecting our outcomes of interest. Inter alia, principals were asked about the size and history of the school, facilities, attrition and teacher quality. We match the twelfth grade school level data with the administrative data using the following combination of information: high school attended, gender, school and national exam scores in each subject examined at the national level, graduation year, year of birth and speciality chosen at the beginning of the twelfth grade. We exclude students who had at least one missing value in those entries. The matching between the dataset provided by the Ministry of Education and the school datasets was very satisfactory [18] providing us with a complete senior high school performance history for 45.746 students which is our sample size.

3. We obtained average household income information for 2009 for every postcode in the country from the Ministry of Economy and Finance. We employ this as a proxy for neighborhood income.

4. We obtained postcode data on urban density information from the Ministry of Internal Affairs. Urban areas are those with more that 20,000 inhabitants.

5. We obtained the Labor Force Survey data for the year 2003 from the National

[14]We exclude from the analysis schools that had at least one year school cohort size smaller than ten students because these small schools may be atypical in some dimensions. Results including those schools are very similar. Contact authors for further results.

[15]Experimental schools are public schools where admission in these schools is based on a randomised lottery.

[16]Which are public schools but lessons take place in the evening targeting employed students.

[17]University cut-offs differ for students graduating from evening schools compared to any other type of school.

[18] 92 % of students matched because of missing values either in the school level data or the administrative Ministry level data.

Statistical Authority. We use quarterly data to create a variable that maps college occupations into annual earnings [19]. We do that if people's reported education is in the same field as their actual occupation in 2003. Respondents report their occupation with high precision [20]. The earnings data are grouped into ten bins indicating the ten national deciles with the highest frequency. We use the lowest bound of each bin [21] to construct a variable that measures minimum expected annual earnings from each occupation.

Every school follows the same curriculum and students are assigned to public schools based on a school district system. This school district system assigns students to schools based on geographical distance. Students are alphabetically assigned to classes in tenth grade and then they do not change class throughout senior high school. Moreover, teachers are allocated to public schools based on geographical criteria and no quality criteria are taken into consideration in the process. Figure 1 presents the geographic position of each school included in the sample. The density of the school population in Athens is 32 % thus many of the schools in our sample are located in Athens[22] .

Table 1 presents descriptive statistics about the available variables in the sample in the twelfth grade. The variable "internal migration" takes the value one if the district of University Department the student is admitted to is different from the district of residence; the latter being proxied by the school district. Moreover, the variable "early enrolment" takes the value of one if the student enrols in the first grade before the age of six [23]. Interestingly, 82 % of the students on average get admitted to at least one University Department. Given that there are no fixed cut-offs, if there is not much demand for a particular University Department the cut-off grade in that year is very low.

Table 2 reports the mean characteristics of the schools in our sample and the whole school population to investigate if our sample is a representative one. There are some variables for which there is a statistically significant difference between the 134 sample schools and the population of schools and these differences are mainly related to the

[19] We also map college fields to occupations.

[20] 209 classified occupations are reported and respondent have to indicate which one is closest to their actual occupation.

[21] Multiplied by 12 months.

[22] In the 2011 census the population of Athens was 3.089,698 while the population of Greece was 10.815.197. Source: National Statistical Authority, 2011 census.

[23] According to the law, this happens if the student is born in the first quarter of the calendar year.

sampling methods that we use [24]. So the sample may not be fully representative of national responses, but it looks pretty similar nonetheless.

3.4 Test Scores

The prior performance measure used in this analysis is based on students' overall school performance in the tenth grade. Each subject's overall grade is a weighted average of the school final exam result and the performance of the student during the school year. We calculate the rank of the student within his school in the tenth grade using his overall grades in thirteen subjects. Teachers receive guidance on how to mark students' exams in the tenth grade and test scores are not curved.

Given the prior performance, we map the effect of the treatment on a composite outcome variable. Our main outcome variable is the relative average rank a student achieves in the five core education subjects measured at the end of twelfth grade. Before 2005 these five subjects were examined at a national level in the twelfth grade. From 2005 onwards, two subjects are examined at a national level whereas the other three subjects are examined at a school level. This change in the number of subjects examined at a national and school level happens at the same year as the abolition of feedback. We do various robustness checks to examine if the results change when we include the rank in each subject separately, the average rank in those subjects examined at the national level or the average rank in the five core education subjects. The test scores in these five subjects is the most important determinant for the calculation of the high school graduation grade under both regimes. In the main analysis, we do not include the elective subjects (or track subjects)[25] in the outcome variable because students choose their electives based on endogenous criteria ie.their perceived differential ability. We show robustness checks later that the results remain unchanged when the Elective subjects are taken into account. Moreover, the subjects included in the outcome variable are compulsory for everyone and test different skills.

On top of the core education exams and the speciality or track subjects, there are compulsory within school exams in three subjects (Sociology, Religion course and Mod-

[24]i.e. the relative percentage of schools in Athens for which we collected data is higher than the relative percentage of schools in Athens. Furthermore, our sample contains 5 % fewer private schools than the population.

[25] Depending on the track students choose in the twelfth grade, they sit national exams in four compulsory subjects within the Track. These four subjects differ from the one Track to the other. The Tracks are: Classics, Exact Science and Information Technology.

ern Greek Literature) in both eleventh and twelfth grades and in both regimes. Students take school exams in these subjects in the eleventh grade and they receive their own score only. In the twelfth grade students are examined again on these subjects without having received any social comparison information in these three subjects before. We call these subjects "non-incentivized", because students' performance in these subjects is not taken into account in the calculation of the university admission grade in any of the regimes. We use these exams as the main counterfactual group.

In this analysis, we use rank measures instead of absolute scores for a couple of reasons. First, using the percentile rank in the tenth grade allows us to do comparisons across cohorts and across schools. Notice that we do not observe the different feedback policies in the same year so we need to compare students who are exposed to different peer groups and teachers. Second, a given twelfth grade national exam score does not represent the same ability level in different years and it is important to make sure that students of the same ability obtain the same relative rank in different years. If the difficulty of the exam changes from the one year to another, then the mean test score changes and any comparison of students' absolute scores across cohorts would be problematic. Furthermore, the mean test scores may be different for different subjects, thus using the percentile rank also allows comparisons across subjects. Also note that any school grade inflation possibly taking place in the tenth grade is not affecting our prior performance measure. Grade inflation would make the teacher more lenient in the overall grading procedure, which implies that the ranking of the students remain unaffected. The national exams in twelfth grade are externally graded, so the teacher in the school has no way to affect students' exam final scores. Furthermore, the national exam procedure does not receive any grade curving.

4 Empirical Strategy

This section identifies the effect of relative performance information on students' senior year exam performance. First, we define the rank measures that we use. Second, we identify if there is an effect. Since we use as an outcome variable the rank in the twelfth grade, the effect is -if anything- of a distributional nature. Then, we discuss the empirical method in order to identify the effect of feedback on students' senior year relative performance.

4.1 Calculation of the rank

In order to calculate the relative rank of the student within his school in the tenth grade, we use the following normalization in order to allow comparisons across schools and cohorts:

$$Rank_{10isc} = \frac{n_{isc}-1}{N_{sc}-1}$$

where n_{isc} is the ordinal rank position of student i within school s in cohort c in tenth grade [26] and is increasing in GPA and N_{sc} is the school cohort size of school s in cohort c. The higher the $Rank_{10isc}$, the higher the rank position of student i in tenth grade in his school s and cohort c. Moreover $Rank_{10isc}$ is bounded between between 0 and 1, with the lowest rank pupil in each school having $R_{10isc} = 0$. For example, in a school consisting of 100 students ($N_{sc} = 100$), the student with the fifth highest GPA ($n_{isc} = 95$) will have $Rank_{10isc} = 0.95$ while the student with the first lowest GPA will have ($n_{isc} = 5$) so his rank will become $Rank_{10isc} = 0.05$.

The ranks of the student within his school in the twelfth grade and nationwide are calculated using the following normalisations:

$$Rank - school_{12isc} = \frac{k_{isc}-1}{K_{cs}-1}$$
$$Rank - nationwide_{12ic} = \frac{r_{ic}-1}{R_{c}-1}$$

Where k_{isc} is the ordinal rank position of student i in school s in cohort c in twelfth grade and is increasing in the national exam grade. K_{cs} is the cohort size c in school s. The $Rank - school_{12isc}$ is projected into the [0,1] interval and the lowest rank pupil in each school cohort has $Rank - school_{12isc}$=0. Notice that there are five exams/subjects, so we first find the ordinal rank of the student based on the average in the five scores, and then we normalise it using the above formula . $Rank - nationwide_{12ic}$ is calculated in a similar way but is irrespective of the school the student attends. So both $Rank - school_{12isc}$ and $Rank - nationwide_{12ic}$ are calculated based on the twelfth grade national exams in the incentivized subjects but they measure relative performance in the school and the country respectively. For example, in a cohort with 50,061 students (R_c=50,061), the student with the tenth highest twelfth grade national exam score (r_{ic}=50,051) will have a national rank of $Rank - nationwide_{12ic}$=0.999. If the same student has 78 schoolmates (K_{cs}=79) and he has the second highest score

[26] Based on the average of the thirteen subjects, ie.the tenth grade GPA.

within his school in that cohort (k_{isc}=77), then the school rank of this student becomes $Rank - school_{12isc}$=0.974.

4.2 Identifying the effect

Figure 3 shows the fitted values of the twelfth grade rank nationwide for each percentile of prior performance. We observe that the fitted regression line for the feedback period is steeper than the non-feedback one, implying that feedback has a positive effect on the better students and a negative effect on the students in the lower part of the ability distribution. Thus, the better students are more likely to end up higher in the twelfth grade rank distribution when feedback is provided. The opposite holds for the worse students who are more likely to end up lower in the twelfth grade rank distribution when they are aware of their previous relative performance.

Figure 4 shows the average rank nationwide that each performance group achieves in the twelfth grade exams, conditional on students' prior performance. Cohorts up to 2005 have received the social comparison information. We observe that the lines are parallel in the pre-treatment period (cohorts 2003,2004 and 2005). This means that the time trends for each quintile of prior performance are following a similar pattern from year to year. There are no evidence of differential pre-treatment trends for the different performance groups. Identification is achieved through a difference approach for each prior performance group. The 2006 cohort is the first affected cohort by the abolition of the rank information provision. We observe that from 2005 to 2006 the slopes of the time trends change, meaning that the treatment affected considerably students in all performance groups except of the middle quintile, which remain unchanged. Another important observation here, is that the slopes remain relatively stable after 2006, which is the first affected cohort. So the change in the slope of the time trends between 2005 and 2006 can be attributed to the abolition of the relative performance information. We produce this figure using students' rank nationwide (Figure 4) and rank within the school (Figure 5). There rank measures are derived using the average rank in the core eduction subjects.

4.3 Method

Here, we quantify the effect of feedback provision on future performance by adopting two complementary strategies.

First, we use the following specification to estimate the effect of feedback information on students' later rank, conditional on their prior performance.

$$Rank - nationwide_{12ic} = \alpha + \beta Feedback_c * Quintiles_{10isc} + \lambda Quintiles_{10isc}$$
$$+\psi Feedback_c + X'\gamma + \psi_c + \phi_s + \epsilon_{ic} \text{ (1a)}$$

$$Rank - school_{12isc} = \mu + \delta Feedback_c * Quintiles_{10isc} + \kappa Quintiles_{10isc}$$
$$+\xi Feedback_c + X'\zeta + \theta_c + \omega_{isc} \text{ (1b)}$$

where $Quintiles_{10isc}$ is a dummy variable that takes the value of one if the student is in the corresponding quintile based on his tenth grade performance in his school. Moreover, $Feedback_c$ is a dummy variable equal to one if the student takes the eleventh grade national exam ie. if the graduation year is smaller than 2006 (feedback regime). The parameter of interest β (δ) measures the effect of feedback on student's rank nationwide (within his school) in the subsequent year, conditional on tenth grade performance. In some specifications, we control for unobserved time and school invariant factors that may affect last year's rank using time and school fixed effects. Specification (1b) exploits within school variation, thus we use (1a) without the school fixed effects when we are interested in exploiting across schools time invariant variation.

In addition to the first strategy, we now use the following difference specification to find the effect of feedback on each decile of students' twelfth grade performance. We run the following specifications for each decile of tenth grade performance $\theta \in [0, 1]$:

$$Rank - nationwide_{12ic\theta} = \delta_\theta + \alpha_\theta X_{i\theta} + \beta_\theta D_{\mathrm{c}} + \psi_c + \epsilon_{ic\theta} \text{ (2a)}$$
$$Rank - school_{12isc\theta} = \omega_\theta + \alpha_\theta X_{ic\theta} + \gamma_\theta D_{\mathrm{c}} + \theta_c + u_{isc\theta} \text{ (2b)}$$

where δ_θ captures a performance group-specific fixed effect. D_{c} is a feedback dummy that takes the value one in the feedback regime and it takes the value zero in the non-feedback regime. The parameter of interest β is estimated separately for each one of the ten deciles, including clusters at the school level. A similar regression across all decile groups gives the pooled OLS estimator of β_θ which is exactly zero because as we explained before, the provision of feedback has a zero average effect. A negative coefficient of β_θ (γ_θ) implies that feedback induces a deterioration in the rank nationwide (within his school) for students at this decile.

5 Main Results

5.1 Effect on performance

Main OLS results are reported in Table 4. The first column corresponds to the basic specification (1a) without school and year fixed effects. The dummy for the third tenth grade quintile is omitted as a point of comparison. This shows that when feedback is provided, a student who is at the top quintile in the country has a 0.042 percentile rank gain in his twelfth grade national exam performance compared to a student who is at the median quintile, ceteris paribus. Similarly, a student who receives feedback and is at the bottom quintile has a 0.088 percentile rank loss in his twelfth grade national performance compared to a student at the median quintile. In columns 2 and 3, we see that the results of column 1 are robust to conditioning out unobserved heterogeneity across schools and years respectively. Adding school and year fixed effects slightly change the coefficients estimates, which remain statistically significant at an 1 % significance level. In all specifications, we control for a set of pupil characteristics and we cluster the standard errors at the school level. These results support the alternative hypothesis of the model discussed earlier.

We now turn to specification (1b) where we exploit the within school variation and results are presented in Table 5. The effect of feedback on students' within school performance in the incentivized subjects is reported in columns (1) and (2) and in the non-icentivized subjects in columns (3) and (4). In the first column, we show that students in the quintiles 5 and 4 (top ones) based on the tenth grade performance benefit from feedback. This gain is associated with 0.045 and 0.040 school percentile ranks respectively compared to the third quintile. Similarly, quintiles 2 and 1 (bottom ones) experience a loss of 0.038 and 0.079 school percentile ranks, when feedback is provided. In column 2 we control for unobserved heterogeneity across years and as we expect; results are similar to Table 4, column 3 when we conditioned out for unobserved heterogeneity across years and schools in the national analysis.

Then, we replicate the same analysis but we now use the school rank in the non-incentivized subjects as the outcome variable. As mentioned before, students take school exams in these subjects in both regimes and grades (eleventh and twelfth). This is a crucial placebo test because if students act as if they receive feedback in these subjects, that would mean that our estimated effect of feedback captures the effect of year unobservables that are not taken out by the year fixed effects. A possible explanation in that case, that would still facilitate our interpretation would be that students might react to feedback by studying more or less for the school instead of the national exams. But still, we would not be able to allay the concern that our estimated effect captures only the effect of feedback and not something else. In columns 3 and 4, we find that the coefficients are not statistically significant and there is no evidence that the provision of feedback affects students' performance in these subjects. What is important here is that students do not receive any social comparison information regarding the non-incentivized subjects neither in the feedback regime nor in the non-feedback regime. These findings are in line with our hypothesis that students change their effort choice and thus their next year performance due to receiving information about their relative performance, when feedback is provided.

We then run specification (2a) and in Figure 6 we plot the β_θ coefficients of the rank nationwide and the associated 95 % confidence interval. We observe that receiving information about someone's relative performance has a negative effect to students below the forty-fifth percentile and a positive effect to students above it. At the highest two deciles, the curve is slightly decreasing implying that there is a ceiling effect. In other words, there is some upper bound on how much improvement can feedback provision bring for the most able students. Thus, sitting similar exams prior to university admission high stake exams improves (decreases) the relative rank nationwide of the high (low) achieving students by up to 5 (8) percentiles. In Figure 7, we report γ_θ coefficients and the associated 95 % confidence interval. The estimated treatment effects on the rank within the school are very similar to the ones found before in Figure 6. This happens because the school sample that we use is a representative one in terms of many observed characteristics and so someone's rank nationwide might not differ a lot from his rank within his school. Figure 8 plots the treatment effect coefficients for the non-incentivized subjects that we use as the main non-treated subjects and we explained before. In line with Table 5, we find no evidence that the provision of feedback affects students' performance in these subjects.

We then standardise the twelfth grade scores in each year and school so that is has a zero mean and a standard deviation of one. Then we run a specification similar to (2b) but the outcome variable is the twelfth grade standardised score of student i in school s in cohort c in each decile θ. We run this regression for each decile of tenth grade performance and we plot the coefficient of the feedback dummy D_c. The treatment effects line for each decile of prior performance is presented in Figure 9. There, the gain for students above the 40th percentile is up to 0.15 standard deviations while the performance of students who are below the 40th percentile drops by up to 0.3 standard deviations.

Someone could argue here that students can accurately place themselves within their class, even if they are not explicitly informed about their rank. This is very likely to occur due to repeated interactions among classmates throughout the high school. However, here students receive new information that is broader than the one they can collect on their own. Consider the within school rank. Students receive information about how well they did within their school. In Figure 11, we report the treatment lines for students in schools of different capacity in the eleventh grade. We make four broad categorisations. First, we consider schools with only one class where it is likely that students already know their relative performance standing and the social comparison information has no extra value (Panel A). Nevertheless, in a school with only two classes students might know their relative performance in their class but not in the school cohort as it contains two classes. Thus, we see that there is a small positive feedback effect on students who are above the 40th percentile and a small negative effect on those below it (Panel B). Additionally, the treatment lines become steeper when we consider schools with three classes (Panel C). In this case, the information given is much broader that the information that students can collect from daily social interaction with their classmates. This is even more pronounced when we look at students in schools with more than three classes (Panel D). Summary statistics about the capacity of schools in our sample are presented in Table 6. Figure 10 shows that the effect of feedback depends on whether the additional information is actually informative about someone's relative performance.

That could possibly allay the concern that the eleventh grade national exam might provide students with experience or training instead of information about their relative performance. School exams in the eleventh grade have the same format as national exams in the eleventh grade and the past papers are available in both regimes. So students practice on the past papers' questions and they are aware of the structure and the types of questions in both cases. If students were equipped with experience from sitting the eleventh grade national exams, then the experience or training effect would not vary by the size of the school. In other words, if that was the mechanism then students in small schools would have no reasons to react differently than students in regular schools.

5.2 Gender and Track

Next, we turn into the gender analysis. As literature on evaluating social programs has shown, individuals respond differently to the same policy (Heckman 2001). To test whether boys react differently than girls to the provision of feedback, we estimate the following regression:

$$Rank - nationwide_{ic} = \delta + \beta Feedback_c * Female_i + \kappa Feedback_c$$
$$+\lambda Female_i + \alpha X_i + \mu_t + \epsilon_{ic} \text{ (3a)}$$

$$Rank - school_{isc} = \delta + \beta Feedback_c * Female_i + \kappa Feedback_c$$
$$+\lambda Female_i + \alpha X_i + \mu_t + \epsilon_{isc} \text{ (3b)}$$

where X_i includes the tenth grade GPA performance, a dummy for early enrollment in school and dummies for the speciality chosen in the twelfth grade. OLS results are shown in Table 7. Although girls outperform boys, girls end up in a lower later rank on average when feedback is provided. This is the case when we consider both; the rank nationwide and the rank within their school. [27] Running specification (2b) [28] for boys and girls separately, produces Figure 11 that presents the treatment lines for boys (on the left) and girls (on the right).

For both genders, the effect of feedback is positive for high achieving students and negative for low achieving students. We make two important points here: First, the average effect of feedback on boys' last year rank is positive and on girls' is negative as shown by the horizontal line which is generated by a regression across all deciles (Figure 11). Second, the effects of feedback are more pronounced for women. As indicated by the steeper treatment line in Figure 11, girls exhibit greater sensitivity to knowing how well they do compared to their school or cohort peers.

Our evidence are consistent with the literature supporting differential gender effect to feedback with females responding more to additional information. McCarty 1986 in an experimental context, shows that women may react differently than men in the absence of feedback information because of different levels of self-confidence. Using an experimental context too, Franz et al. 2009 argue that women never have the same level of self-confidence as men because women expect less of themselves than men do.

[27] In Table 7, if we include school fixed effects in columns (3) and (4), we account for heterogeneity across schools and the coefficient estimates become the same as in columns(1) and (2).

[28] (2a) gives almost identical results as (2b) for both genders.

Then, we disaggregate the analysis at the Track level. There are three Tracks: Classics, Exact Science and Information Technology and students have to take four exams within each track [29]. In Figure 14, we run specification (2a) separately for each track. The smallest average effect is observed for students in the Science Track whereas the treatment curve is steeper for students in the Information Technology Track rather than in Classics.

5.3 Long term outcomes

In this section we examine the effect of feedback provision on students' long term outcomes. We have already motivated the discussion regarding the reasons a student would choose to resit the national exams for university admission. We use binary response models to examine whether the provision of feedback affects the decision to retake the exam. In Table 8, we observe that a significant percentage out of the cohort population repeats the exams one year after graduation from senior high school[30].

We define as "misplacement" the difference between the tenth grade rank within the school each student gets and the rank nationwide in the twelfth grade. Thus, the misplacement variable is bounded between minus one and one. Students with larger differences between the tenth and the twelfth grade ranks would have a large change in their relative performance. The misplacement variable takes the value zero for students where their twelfth grade rank happens to correspond exactly to the tenth grade rank. But it can also take positive (negative) values if the student achieves a better (worse) relative performance in the tenth grade relative to the twelfth.

In order to examine if feedback provision affects someone's decision to retake the national exams through the misplacement effect we run the following specification:

$$Retake_{i,t+1,s,d} = a + X'_{itsd}\gamma + \delta Misplacement_{itsd}Feedback_t + \beta Feedback_t$$

$$+\omega Misplacement_{itsd} + \zeta Z_{td} + \xi_s + \omega_t + \epsilon_{itsd}$$

The decision to retake the national exam one year after graduation depends also on the opportunity cost of the student. Thus, we control for the unemployment rate in each year t and district d of student's residence.

[29] In Classics they take national exams in: Ancient Greek, Latin, Literature and History. In Science the examined subjects are: Mathematics, Physics, Chemistry and Biology and in Information Technology: Computers, Mathematics, Physics and Business Administration.

[30] The number of students retaking the exam is calculated using the Ministry of Education dataset. The data about the labour force capacity are collected from the National Statistical Authority.

Using Linear Probability (LPM), Probit and Logit models we find that when feedback is provided, students with higher misplacement are more likely to repeat the national exams one year after graduation. In Table 10, we interact dummies that capture the magnitude of misplacement with the feedback dummy and we observe that students in the top misplacement quintile (5) are more likely to resit the national exams when feedback is provided. The Top Misplacement Quintile (5) is the most positive one and contains students who get a better rank in the tenth grade compared to the twelfth. In the feedback years, these are the low achieving students. In other words, low (high) achieving students are more (less) likely to resit the national exams when feedback is provided.

Having a particular placement in university admission affects the employment and earnings prospects of an individual. We examine if feedback influences the matching of students to University Departments. We first rank all programs [31] according to their average cutoffs over the seven years period. Each program's cutoff expresses the students' valuation for this particular university department. Highly demanded programs exhibit high cutoffs. Students apply to programs based on preferences, social status and expected earnings. There are 659 programs in total. We estimate the effect of feedback on the difference in the popularity position and rank of the program admitted conditional on tenth grade performance. Figure 15 presents the treatment effect line for the popularity position (on the left) and rank of the program (on the right) admitted. The provision of feedback has a positive effect on the popularity position and rank of the program admitted in the upper half of the prior performance distribution and negative effect on the low half. In particular, high achieving students move up the University popularity ladder by 30 positions which is 0.15 of a standard deviation. Different placements in university admission induce different gains related to the returns to college.

Enrolling into a specific University Department may affect students' career path and their lifetime earnings. We match administrative salary data for each occupation from the Labor Force Survey to each University Department. In particular, we use the 2003 Labour Force Survey to map each college major into the most related occupation and then into the expected annual earnings after graduation (in Euro).[32] We then use these figures as the expected earnings of current students after graduation from the

[31] By program we mean each combination of University Department.

[32] Mean:12,758 with 1,473 standard deviation.

particular program. In Figure 16, we present the effect of feedback on the expected annual earnings, conditional on the tenth grade performance. For students above the 50th percentile, there is an increase in their annual expected earnings by 250 Euros per year, which is equivalent to 0.17 standard deviation. For students below the 50th percentile, the decline in their expected annual earnings corresponds to 0.20 standard deviations.

5.4 Social Mobility

In this section, we examine if the provision of feedback changes the relationship between family income (proxied by neighborhood income) and post-secondary opportunities (indicated by the program the student enrols in). A priori, we might expect that students coming from more advantaged neighborhoods have higher chances of embarking on a better and more prestigious program with higher expected returns than students coming from less advantaged neighborhoods. The neighborhood income is associated with parental income while the program each student manages to get in is associated with the income of destination.

In order to investigate if feedback changes this mapping from parental income to income of destination, we create quintiles based on the neighborhood income and the popularity of the program admitted to. For each quintile of neighborhood income, in Table 13 we report the percentage of students who enrol into each quintile of programs by popularity in the feedback and the non-feedback regime. Then, we calculate the difference between the feedback and the non-feedback percentage. In the last row of Table 13, we vertically add the percentages of students who enrol in any program in order to find the total difference of enrolled students between the feedback and the non-feedback period.

Providing relative performance information might affect differently students with different parental income, as it may be related to other family resources (financial support or social networks). When students from low-income neighbourhoods realise in the eleventh grade that they are highly ranked in a nationwide competition scale, they might react by exerting more effort. In the first column of Table 13, we find descriptive evidence that feedback alters the socio-economic background composition of students who manage to get admitted into the top-ranked programs. More students from low income neighborhoods get admitted to the most prestigious programs in the country with the highest expected earnings after graduation (like engineering and law), when feedback information is provided (2.9% Vs 2.6 %). This implies that feedback information encourages social elevation. And these students are more likely to increase their income of destination after graduating from the University. We also find that more students who belong in the first quintile of neighborhood income get admitted to any program (2.2% more students), when feedback is provided. From the descriptive evidence presented here, feedback has benefited students from low-income neighborhoods by reducing social inequalities and possibly future income inequalities.

5.5 Positive Vs Negative Surprise

In this section, we examine whether students respond to the specific type of feedback that they get. Students might not only compare themselves with their class or school or cohort-mates but they may also compare their own relative performance in different periods in time. Exploiting within school variation in the 134 senior high schools and the fact that we know the whole distribution of scores, we restrict this part of the analysis into the feedback years. If a student receives information that he is in a higher (lower) decile in the eleventh grade than in the tenth grade, then our student receives a positive (negative) shock, that can be translated into a "positive (negative) surprise". Intuitively, students who receive a positive (negative) surprise in the eleventh grade might increase (decrease) their expectations of themselves and exert more (less) effort in the twelfth grade. In order to examine potential effects coming from the surprise they experience in the eleventh grade, we graph the effect on the twelfth grade rank for each combination of percentile ranks in the tenth grade and eleventh grade. That is shown by the heatplot in Figure 18.

The horizontal axis represent the eleventh grade percentile rank of students and the vertical axis represent the tenth grade percentile rank. Different colours express different magnitudes of the treatment effects on the twelfth grade rank. The diagonal starting from zero towards the right upper edge of the box, represents the case of "no value feedback" or in other words those students with their eleventh grade percentile rank equal to the tenth grade percentile rank. The treatment effect is positive (negative) for most students experiencing a positive (negative) surprise.

A concern here is that students might not be aware of their tenth grade percentile rank, especially if they attend a school with more than one classes. However, the analysis here is done for deciles of performance and not for percentiles, allowing students to have priors that do not accurately express their exact tenth grade rank.

5.6 Alternative mechanism: School quality revelation

An alternative mechanism could be that students use the information obtained by the publication of their scores in such a way that they realise the quality of their senior high school[33]. Students who take the eleventh grade national exams suddenly realise their school rank and their national rank and the comparison of the two ranks reveals information about the quality of the school. If a student realises that his national rank is greater than the school rank then his school is of good quality. The opposite if the national rank is lower than the school rank. The realisation of the school quality in the eleventh grade might affect students' choice of effort in the twelfth grade. Thus, we exploit the across schools variation in their quality to identify the effect of feedback on students' rank nationwide.

In Figure 12, we produce the treatment lines separately for students who realise that the school they attend is worse (on the left) and better (on the right) than the average quality school. In Figure 13, we repeat the same exercise and we produce these figures using the standardised national exam score. [34] The average effect for students who realise that they attend a worse than average quality school is negative whereas it is positive for those who realise that they attend a better than average quality school.

Starting with the bottom of the prior performance distribution, we observe that low achieving students in good schools do better that those in lower quality schools. Surprisingly, there is a huge increase in the national rank for the top students in the worse schools and this increase even offsets the increase in the national rank of the top students in good schools. We acknowledge two possible explanations here: First, the better students in the worse schools take the eleventh grade national exams, they receive feedback, they realise that they are actually exceptional in a national scale and thus decide to exert more effort in the next time period. So feedback acts as a motivation boost for these students.

[33] We measure school quality based on the schools' average national exam performance in the twelfth grade from 2003 to 2009. Then we construct a rank measure for school quality that varies from zero to one. The average quality of the schools in our sample is 0.52 (sd 0.21, Min 0.018 and Max 0.985) which means that our school sample is of a representative quality.

[34] Standardised within each year with zero mean and a standard deviation of one.

Second, the realisation of their national rank act as a rude awakening for these students who might initially have a wrong perception about the national competition and about their school's quality. These students might be the top students in their class or school but they now learn that they are left behind. In the next time period, they exert more effort in order to catch up with the national competition.

The fact that the two curves do not follow the same pattern enhances the argument that the results are not driven by experience or practice resulted from sitting the eleventh grade national exam. If students realise the quality of the high school through the eleventh grade national exams, then would all receive the same information and they would not react so differently.

6 Threats to identification

Attrition

In our attempt to evaluate the impact of feedback on different performance groups, the problem of attrition cannot be ignored. If attrition is random and affects different performance groups in a similar way in both regimes, then the estimates remain unbiased. Differential attrition here could arise because students from the lowest percentiles are more likely to drop out from school in comparison to students from the highest percentiles when they realise their relative ability performance. If that is stable over time, it will not affect out feedback estimates. What could bias our estimates, is if differential attrition follows the abolition of feedback. [35] In Figure 17, we observe that attrition rates differ for each quintile but the difference in attrition rates does not change dramatically before or following the abolition of feedback.

Exploiting within school variation, we use the following specification to check for differential attrition that changes with feedback:

$$Drop-out_{12-10isc} = \alpha + \beta Feedback_c * Quintiles_{10isc} + \lambda Quintiles_{10isc}$$

$$+\psi Feedback_c + X'\gamma + \theta_c + \varphi_s + \epsilon_{isc}$$

Table 11 reports OLS results. The attrition rate is larger for the lowest quintile than any other compared to the third quintile when feedback is provided. But most

[35] The first affected cohort for which feedback is abolished is the cohort that was in the twelfth grade in 2006. Thus, this cohort was in the tenth grade in 2004. This is the first cohort that did not sit national exams in the eleventh grade.

importantly, none of the coefficients of interest are statistically significant. This implies that there is differential attrition, but it does not vary with feedback policies.

Robustness checks

In this section, we construct a robustness exercise to complement our main analysis. One concern is that the change in the variation of performance over time might not be caused by the provision of feedback. In other words, we need to rule out the possibility that the better students become worse over time and the worse students become better over time for reasons different than the provision of feedback.

Exploiting the within school variation, we run specification (2b) but without pooling feedback and non-feedback years together. Instead, we just compare every pair of consecutive years in the sample. The only pair of years that we expect to find a differential response of cohorts is 2005-2006 (the year of the reform). For every other pair of years, we expect to find similar cohort behaviour. We present the placebo treatment lines in Figure 13. Panel A compares the cohort 2003 to the 2004, as if feedback was abolished in 2004. We find no evidence that other factors might affect students differently in other years. Panel C corresponds to the actual reform and we observe that the treatment effects are negative for all percentiles below the 50th percentile and positive above it. Regarding any policy anticipation effects, the reform was announced in around December of 2003-2004. We find very small treatment effects in Panel D, which is the first non-treated cohort. Students in the first non-treated cohort might observe how last year's peers of similar tenth grade performance did and use this information to adjust their effort. Again after 2007, the curve is almost flat throughout the ability distribution.

We conduct some other placebo exercises to verify that the effect does not depend on the numbers of subjects examined. In Figure 21, we draw the treatment lines for each subject separately. In Table 12, we calculate the twelfth grade rank based on different subjects. In column (1) we find the effect of feedback on the last year rank that takes into account the Electives or Track subjects on the top of the core education subjects[36] and the results are very similar to those reported so far. In column (2) we take into account the effect of feedback on students' performance in Modern Greek which is a

[36]Students sit national exams in four Elective subjects. So the overall rank in calculated based on nine subjects.

common subject in both regimes and takes a special weight in the calculation of the University admission grade. Notice, that in the non-feedback regime two subjects are examined nationally and three within the school. In column (3) we calculate the last year's rank based on the five subjects in the feedback regime and the two subjects in the non-feedback regime. Results remain very similar. Treatment effect remain positive (negative) for the high (low) achieving students.

7 Conclusion

In this paper, we examined the effects of providing relative performance information on students' short and long term outcomes. We exploit a large scale natural experiment that took place in Greece and thus we conducted a large scale primary data collection process. Using detailed data on students' performance throughout senior high school and school quality data, we examine the effects of receiving information about someone's national and school relative performance. It is human nature to make comparisons, which can affect students' beliefs about own ability and effort decisions. For students above (below) the 50th percentile, we found that feedback information has a positive (negative) effect on their subsequent performance, popularity of program admitted and expected annual earnings.

We outlined two potential mechanisms in this study for why students would react to the provision of feedback. The first one supports that with feedback, students update their belief about their own relative ability and that determines the next period's effort choice, as explained in the theoretical model. Another possible mechanism is that students combine the country and school level information about their ranks that reveals new information about their school quality. Knowing the school quality might provide information to students about the level of the competition over restricted university places. We use these mechanisms to explain our results.

Our findings have important policy implications both in relative and absolute terms.

First, the effects of feedback are positive on the high achieving students and negative on the low achieving students implying that policy makers need to be cautious depending on who they target. These effects concern students' next year performance but also long term outcomes. Feedback provision affects the matching with the university department students are admitted to and consequently their life term earnings. Secondly, girls are more sensitive to feedback and they respond more at both tails of the ability distribution. The relative nature of the above mentioned results restrict the broad implementation of feedback, but makes it very important in a competitive process. Our analysis highlights the importance of rank position on students' scholastic and labour market outcomes and we believe that the rank could be a new factor in the education production function.

Our analysis moves on highlighting the absolute effects of feedback: First, high achieving students in worse schools gain a lot from feedback. Second, the consequence of no feedback is more resitting for high achieving students. This is an important loss of human capital for the society given that the most able students stay out of the university and/or the labour market. Third, we find evidence that feedback encourages students from low-income neighbourhoods. More precisely, more students from low-income regions gain admission to top University Departments when feedback is provided, indicating a potential future drop in income inequality.

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Quarterly Journal of Economics 25*(1), 95–135.

Angrist, J. and V. Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics 114*(114), 533–575.

Angrist, J. D. and V. Lavy (2001). Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools. *Journal of Labor Economics 19*(2), 343–69.

Azmat, G. and N. Iriberri (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics 94*(5), 797–811.

Bandiera, O., V. Larcinese, and I. Rasul (2008). Blissful Ignorance? Evidence from a Natural Experiment on The Effect of Individual Feedback on Performance. *Working Paper*.

Barankay, I. (2012). Evidence from a Randomized Workplace Experiment. *Management Science, Working Paper*.

Brunello, G. and M. Schlotter (2011). Non-Cognitive Skills and Personality Traits: Labour Market Relevance and Their Development in Education and Training Systems. *IZA Working Paper* (5743).

Card, D. and B. A. Krueger (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *The Journal of Political Economy 100*(1), 1–40.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review 102*(6), 2981–3003.

Chetty, R., J. Friedman, and J. Rockoff (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economics Review 104*(9), 2593 2632.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review 101*(5), 1739–1774.

Ertac, S. (2005). Social Comparisons and Optimal Information Revelation:Theory and Experiments,. *mimeo, University of California*.

Franz, T., E. Frick, and K. Hanslits (2009). Gender Differences in Response to Failure Feedback. *Working paper*.

Hannan, R. L., R. Krishnan, and A. H. Newman (2008). The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review 83*(4), 893–913.

Hanushek, E., F. J. Kain, and G. S. Rivkin (2005). Teachers, Schools and Academic Achievement. *Econometrica 73*(2), 417–458.

Heckman, D., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy 109*(4), 673–748.

Hoxby, C. M. (2000a). Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Paper 7867*.

Hoxby, C. M. (2000b). The Effects Of Class Size On Student Achievement: New Evidence From Population Variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

Kane, T., J. Rockoff, and D. Staiger (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review 27*(6), 615–631.

Kautz, T., J. Heckman, R. Diris, B. t. Weel, and L. Borghans (2014). Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success. *OECD Education Working Papers 110*.

Krueger, B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics 114*(114), 497–532.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *Economic Journal, November issue*.

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Economic Journal 122*(559), 208–237.

Lavy, V., O. Silva, and F. Weindhardt (2012). The Good, The Bad and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics 20*(12), 367–414.

McCarty, P. A. (1986). Effects of Feedback on the Self-Confidence of Men and Women. *Journal of Academy of Management 29*(4), 840–847.

Murphy, R. and F. Weinhardt (2014). Top of the class: The Importance of Rank Position. *CEP Discussion Paper No. 1241*.

Rivkin, S., E. Hanushek, and J. Kain (2005). Teachers, Schools, and Academic Achievement. *Econometrica 73*(2), 417–458.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review 94*(2), 247–252.

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics 125*(1), 175–214.

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

Figure 1: Map of schools in the sample

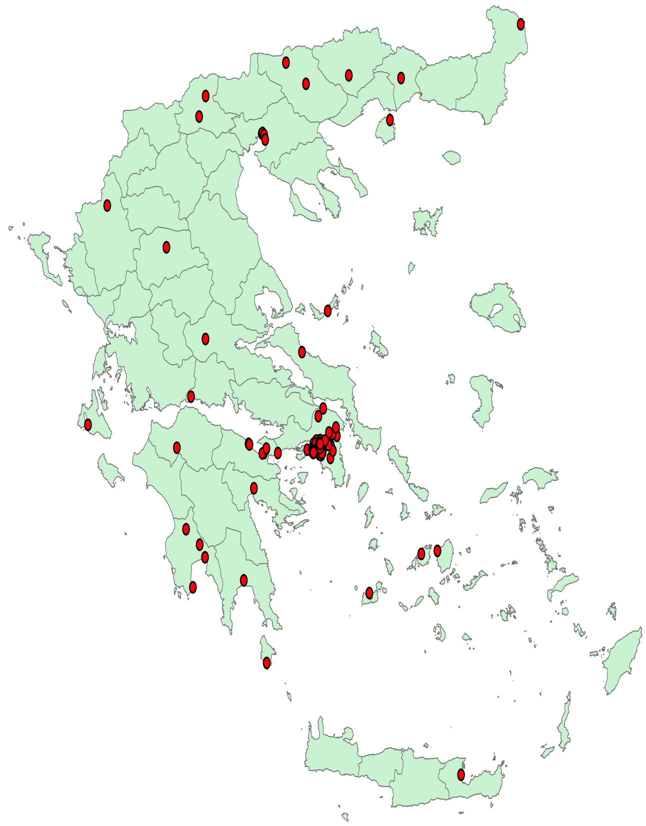


Figure 2: Timing

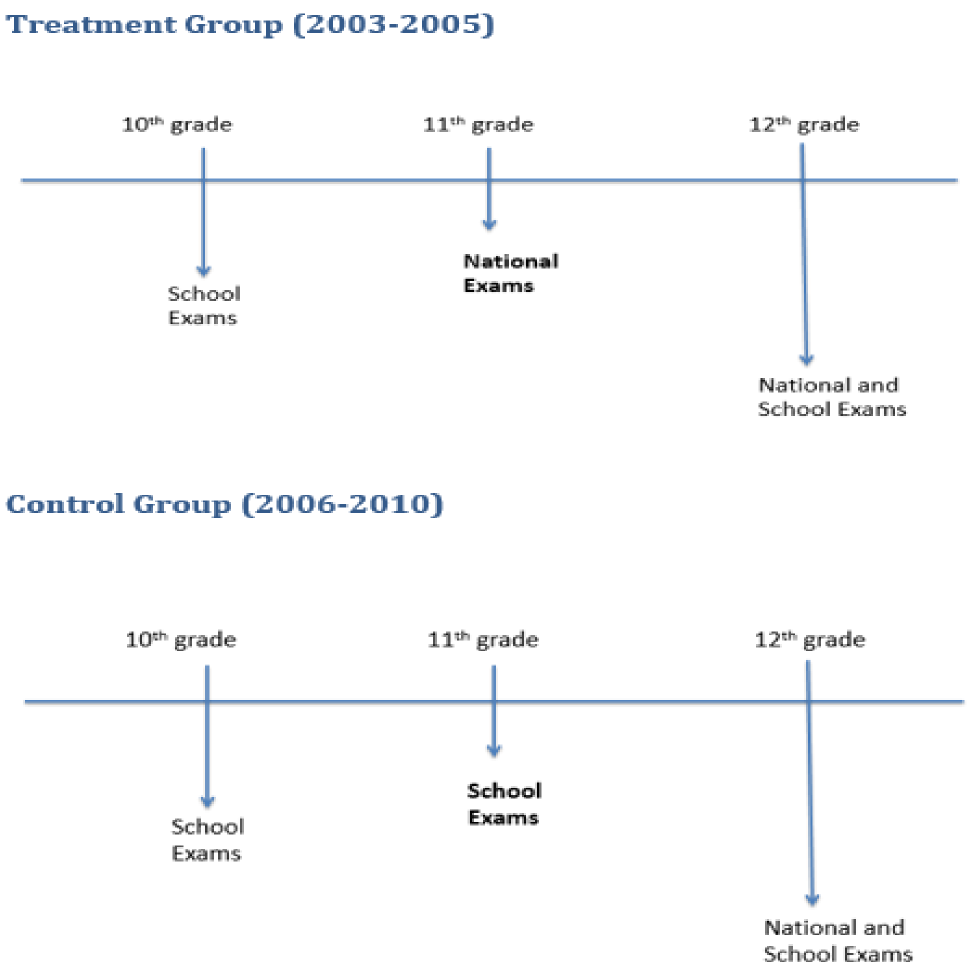
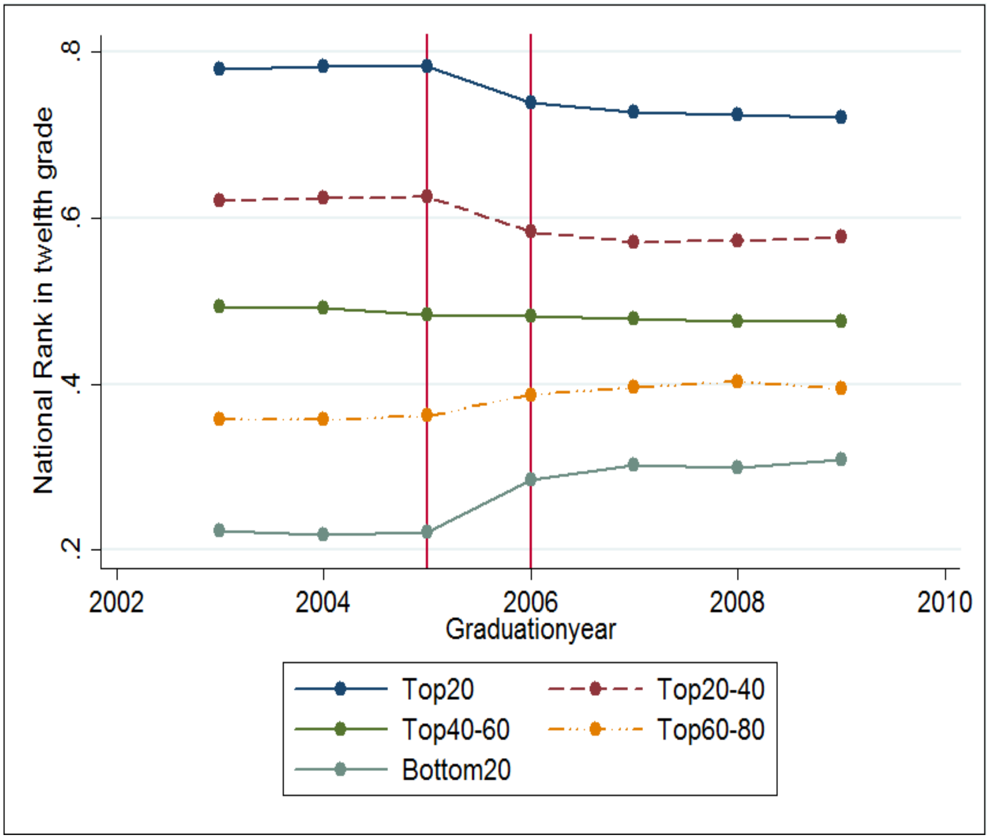
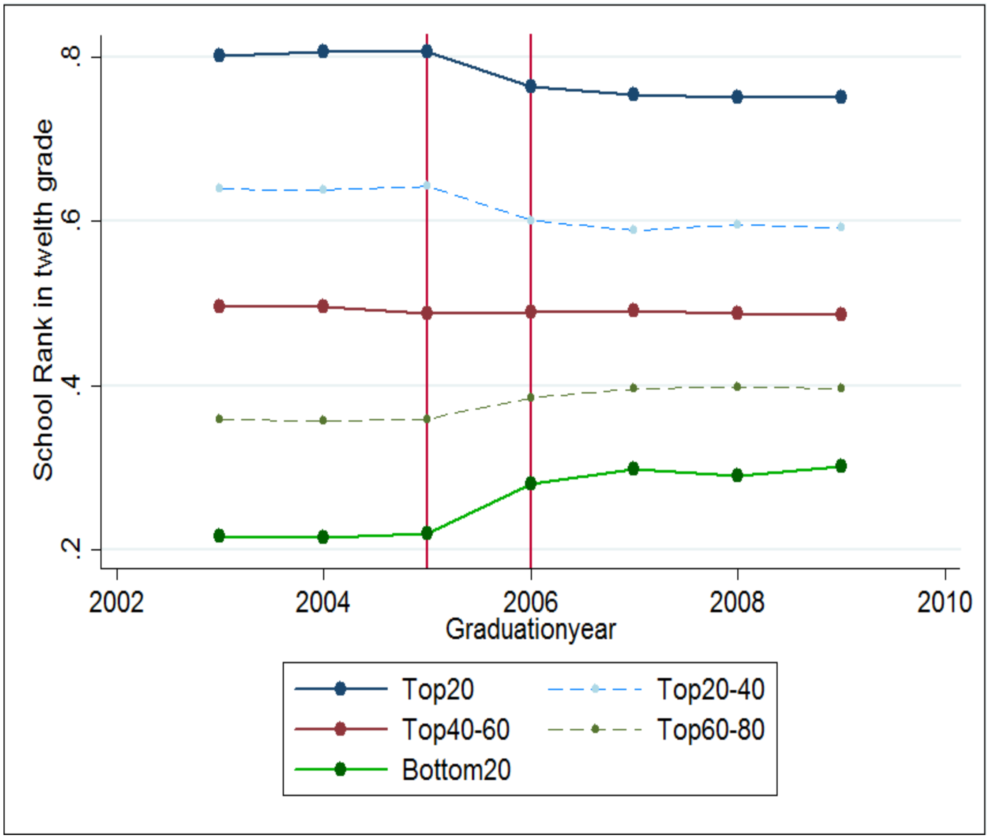


Figure 3: Time trends for twelfth grade rank nationwide



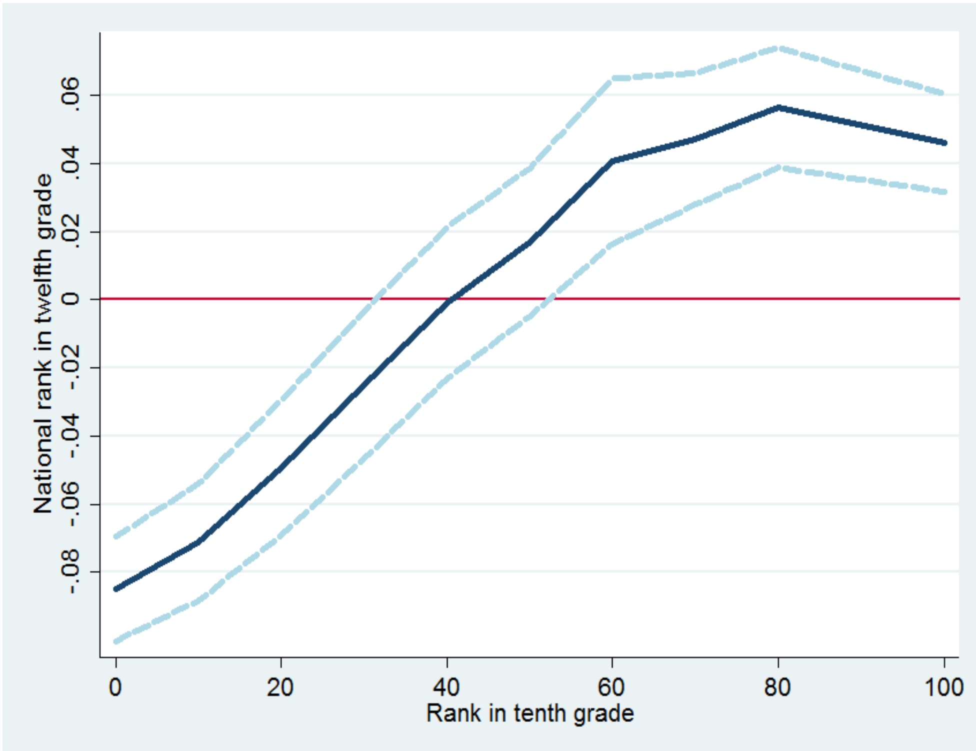
Note: Feedback provision for cohorts 2003-2005.The 2006 cohort is the first one for which feedback is abolished. Outcome variable: The national rank in twelfth grade. The trends correspond to different performance groups based on the tenth grade performance.

Figure 4: Time trends for twelfth grade rank within the school



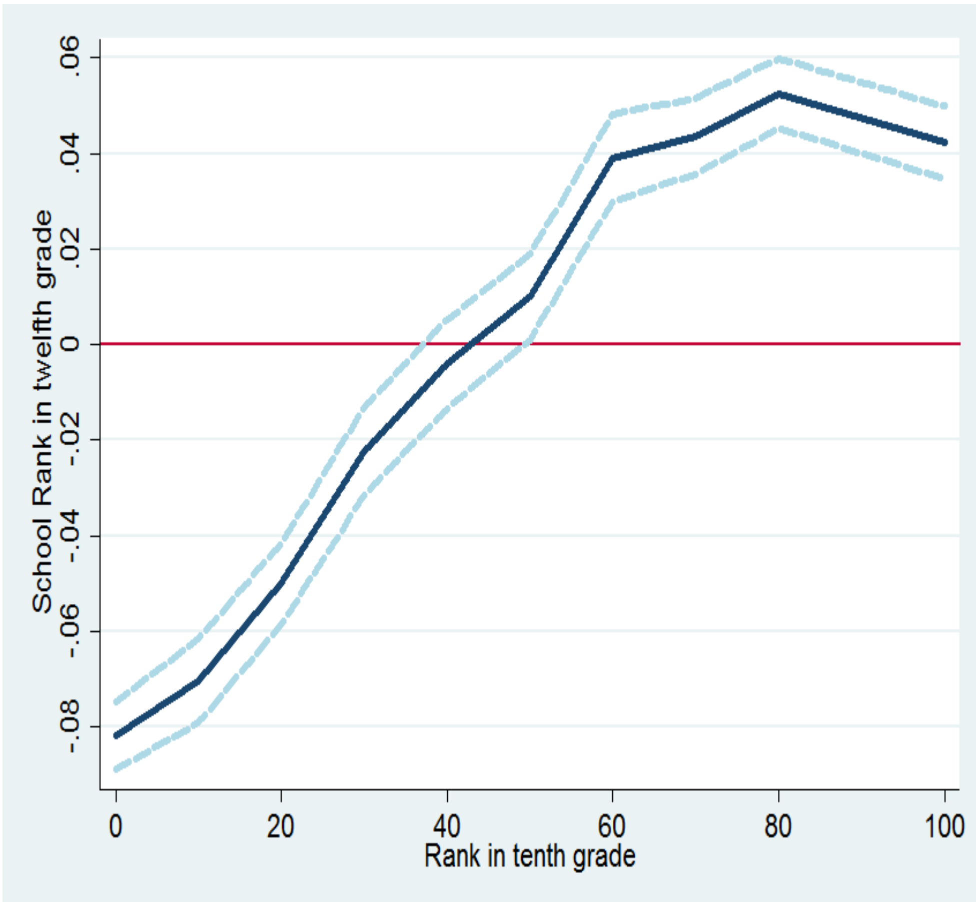
Note: Feedback provision for cohorts 2003-2005.The 2006 cohort is the first one for which feedback is abolished. Outcome variable: The rank in twelfth grade within the school. The trends correspond to different performance groups based on the tenth grade performance.

Figure 5: Treatment effects on the rank nationwide conditional on prior performance



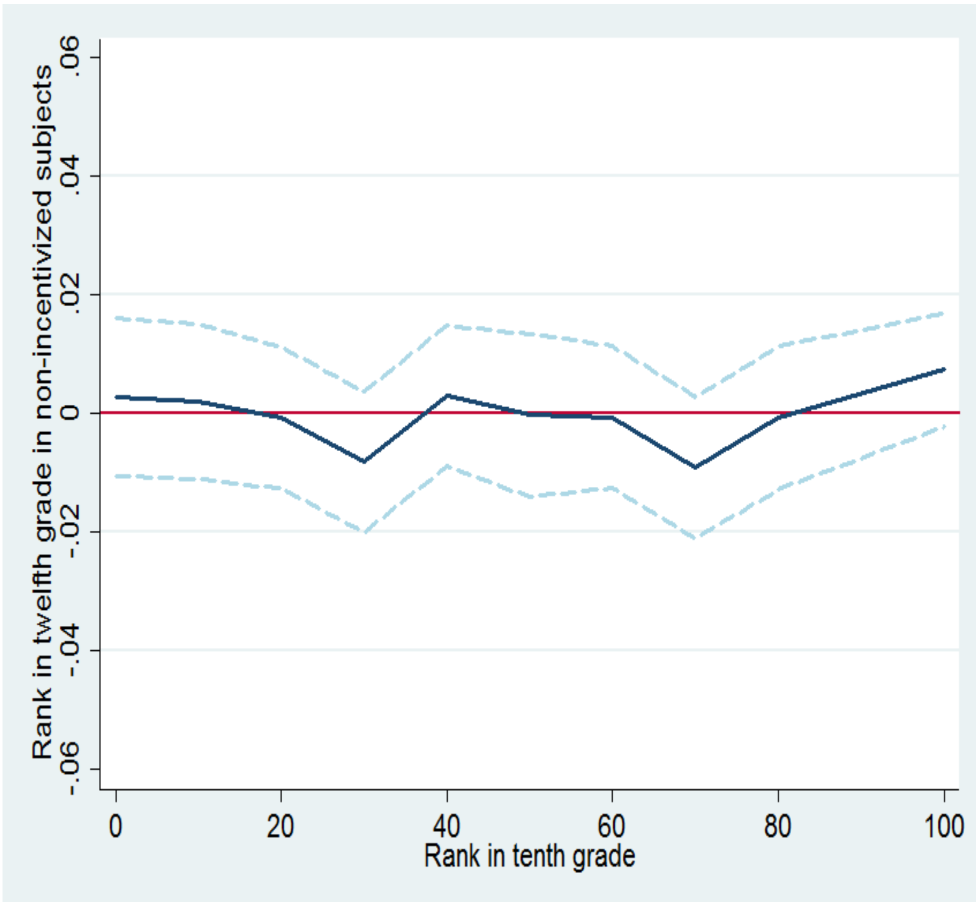
Note: The estimated effect of feedback on the national rank in the twelfth grade at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The national rank is calculated based on the five core educational subjects (incentivized). The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 6: Treatment effects on the rank within the school in incentivized subjects conditional on prior performance



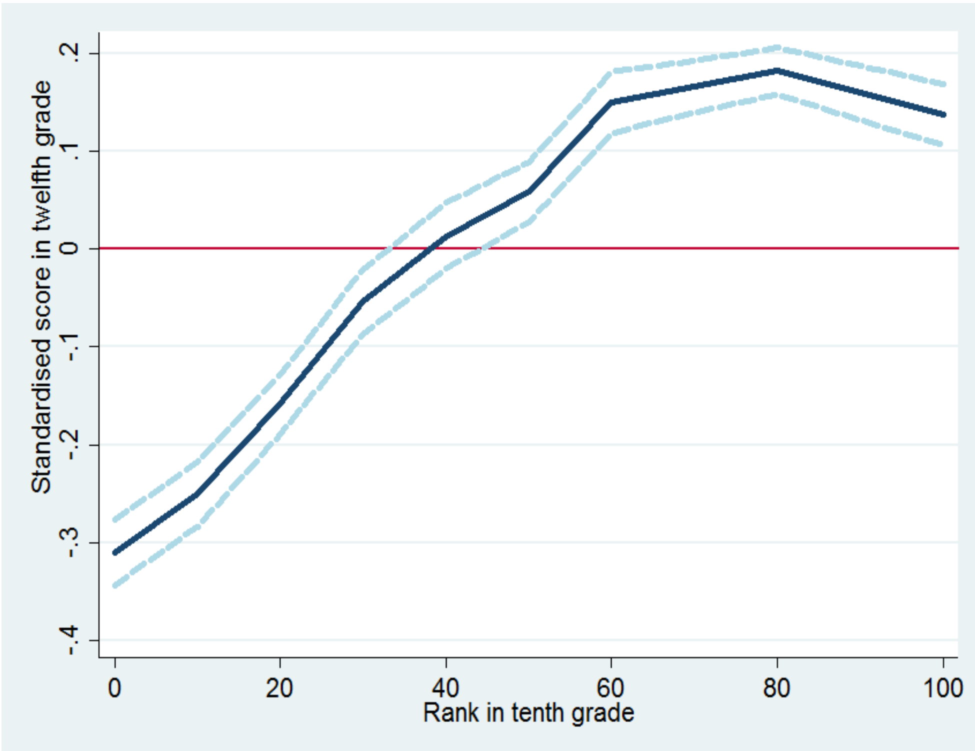
Note: The estimated effect of feedback on the school rank in the twelfth grade at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The school rank is calculated based on the five core educational subjects that students take in the twelfth grade and determine the University admission grade (incentivized subjects). The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 7: Treatment effects on the rank within the school in non-incentivized subjects conditional on prior performance



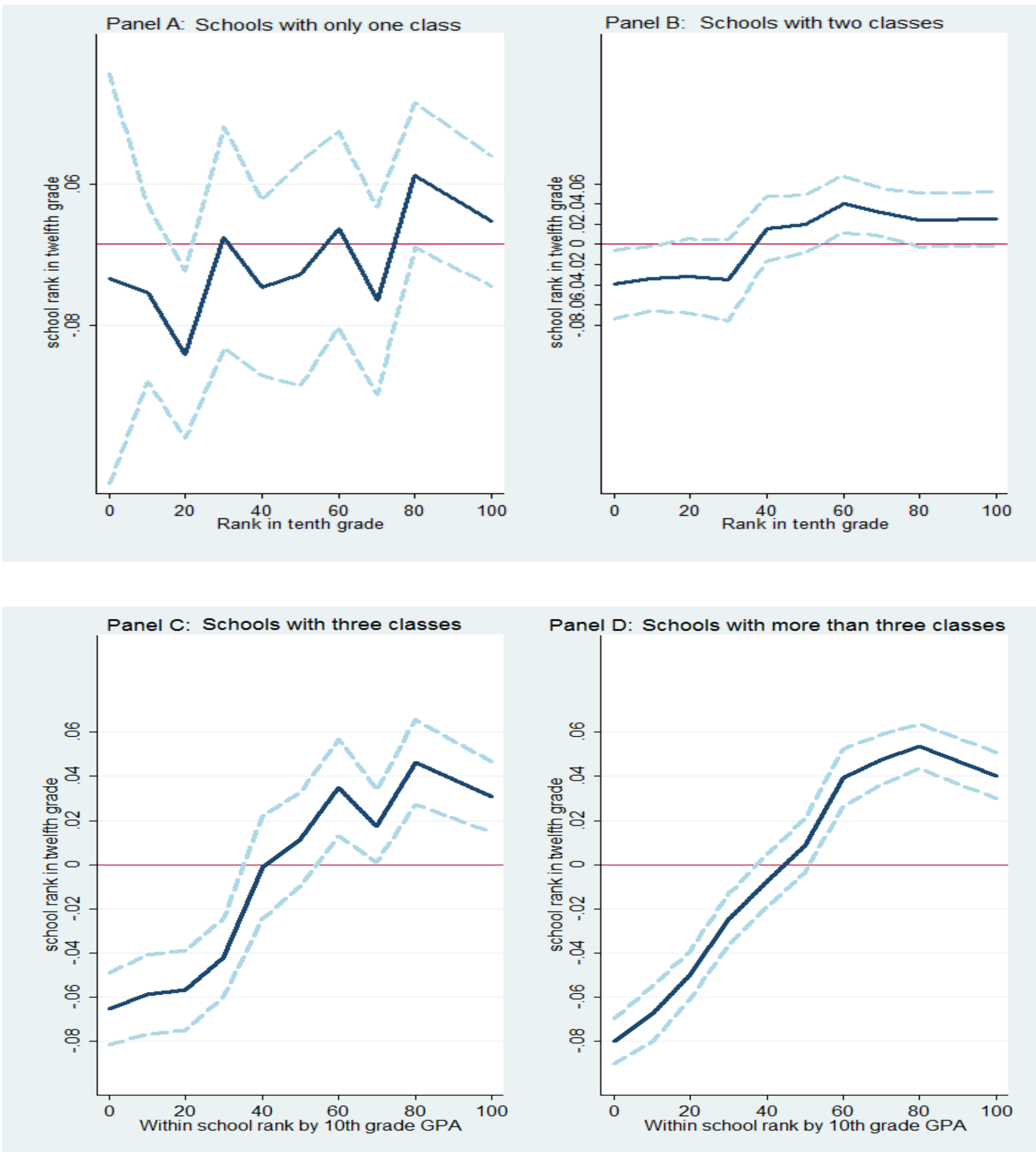
Note: The estimated effect of feedback on the school rank in the school exams at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The school rank in the school exams is calculated based on the three non-incentivised subjects that all students take in the twelfth grade and these subjects are not taken into account in the calculation of the University admission grade. Students never receive social comparison information in these subjects. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 8: Treatment effects on the standardised score conditional on prior performance



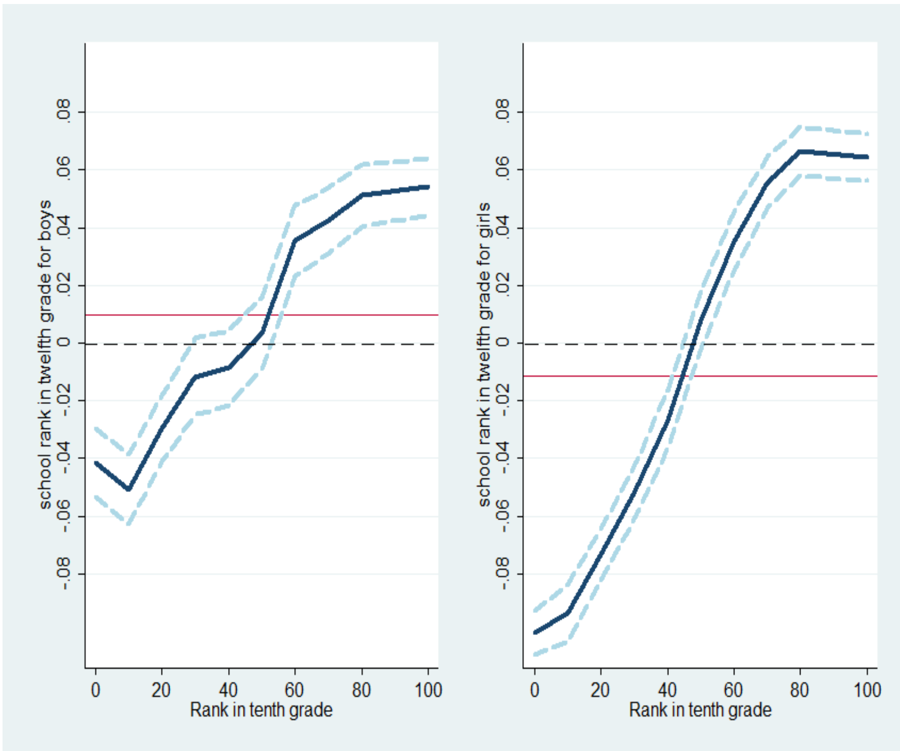
Note: The estimated effect of feedback on the standardised score in the twelfth grade at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The standardised score is calculated based on the five core educational subjects (incentivized). The standardised score has a mean of zero and a standard deviation of one in each year. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 9: Treatment effects on the rank within the school conditional on prior performance for different school cohorts' size



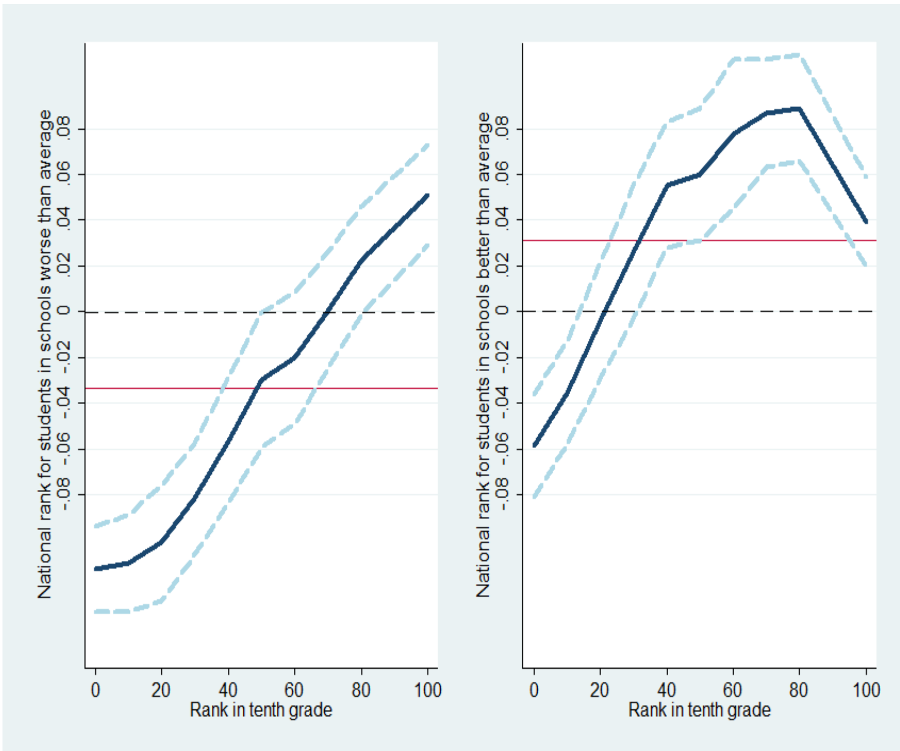
Note: The estimated effect of feedback on the school rank in the twelfth grade by capacity of school at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 10: Treatment effects on the rank within the school by gender conditional on prior performance



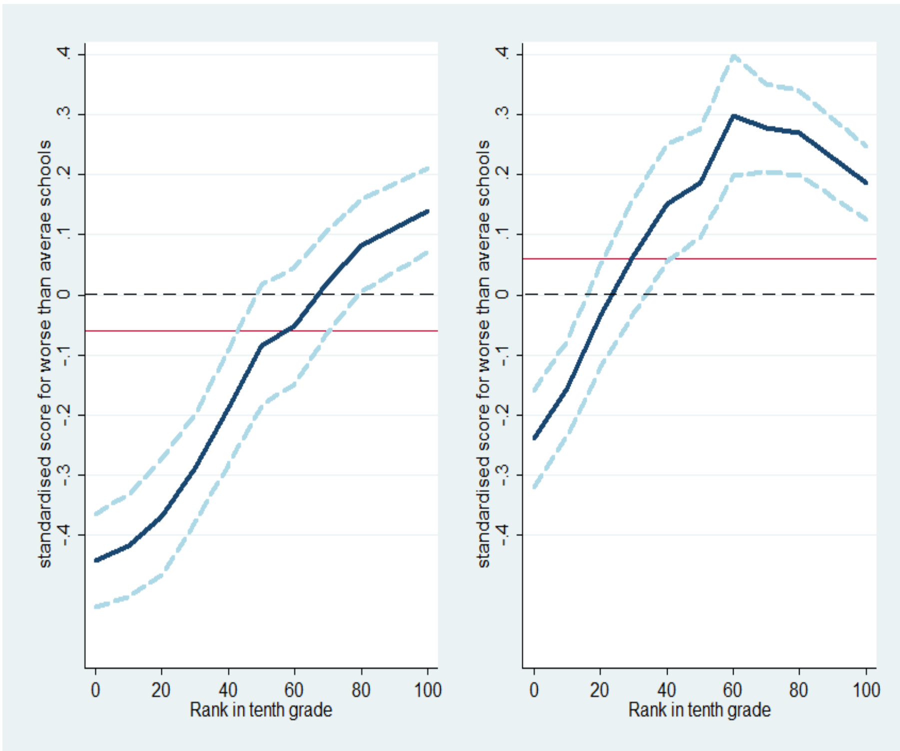
Note: The estimated effect of feedback on the school rank in the twelfth grade by gender at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. Males are depicted on the left and Females on the right. The school rank is calculated based on the five core educational subjects (incentivised). The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 11: Treatment effects on the rank nationwide by school quality conditional on prior performance



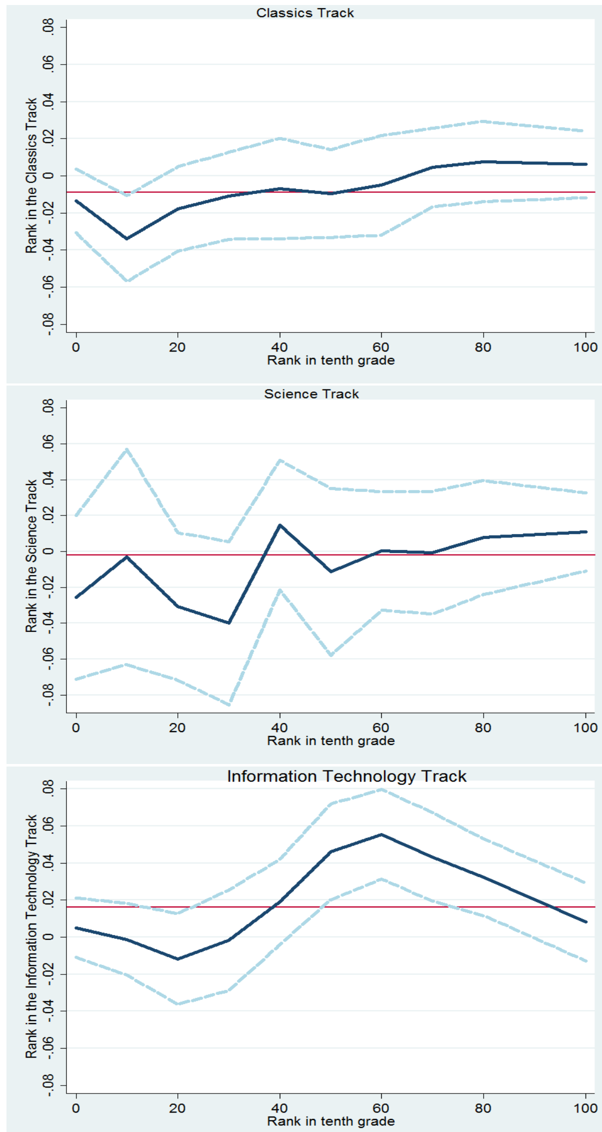
Note: The estimated effect of feedback on the national rank in the twelfth grade by quality of school at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The effect of feedback on students' national rank when they realise they are in schools worse than the average quality school (on the left) and better than the average quality schools (on the right). The national rank is calculated based on the five core educational subjects (incentivised). The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 12: Treatment effects on the standardised score by school quality conditional on prior performance



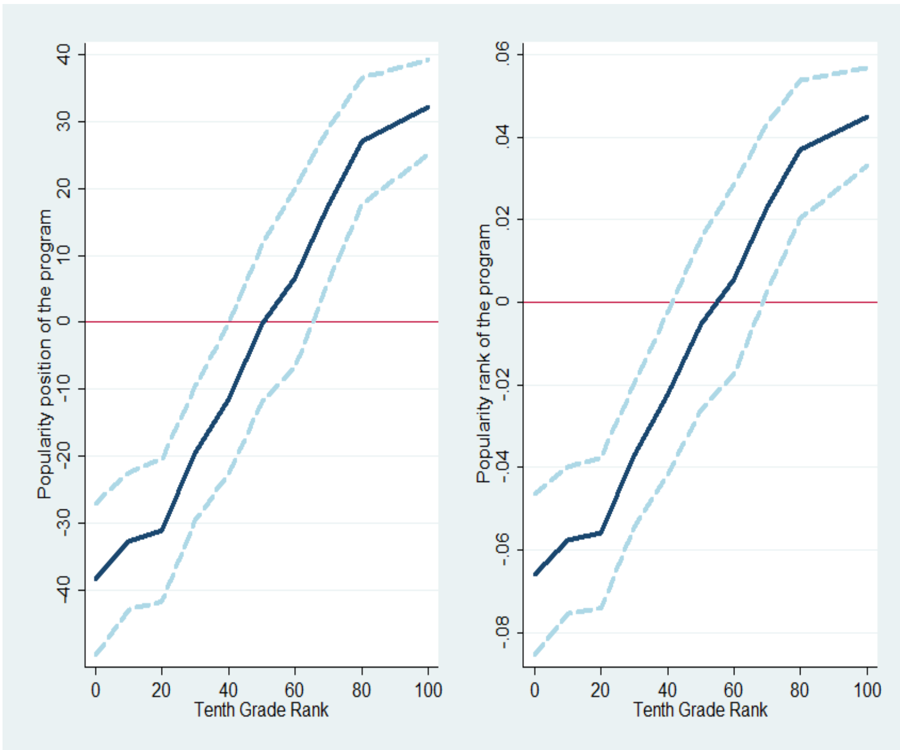
Note: The estimated effect of feedback on the standardised score in the twelfth grade by quality of school at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The effect of feedback on students' standardised score when they realise they are in schools worse than the average quality school (on the left) and better than the average quality schools (on the right). The standardised score is calculated based on the five core educational subjects (incentivized). The standardised score has a mean of zero and a standard deviation of one in each year. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 13: Treatment effects on the rank nationwide by school quality conditional on prior performance



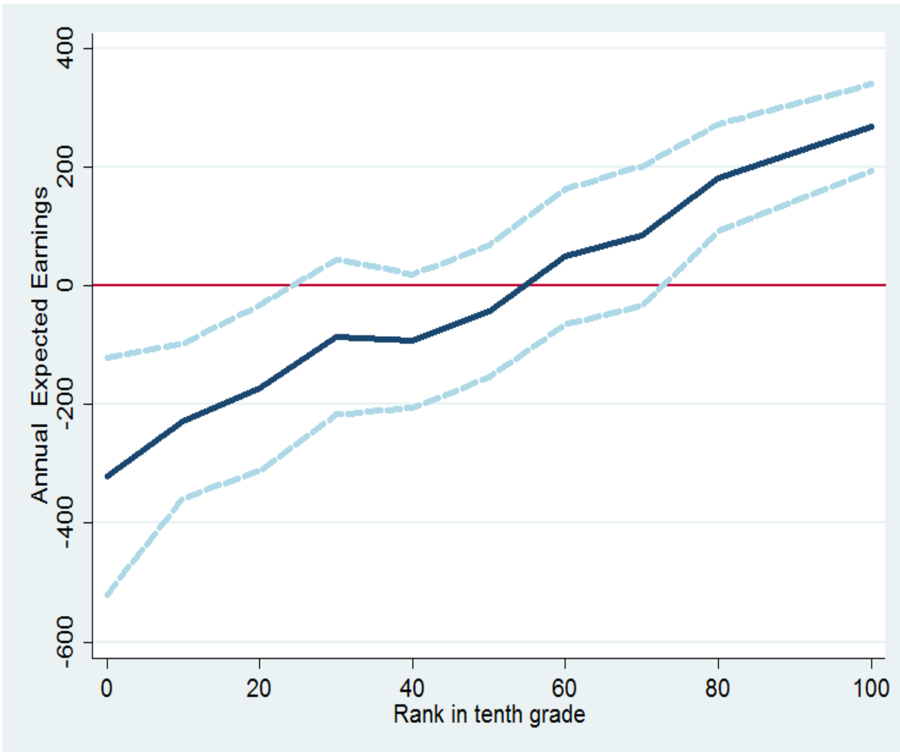
Note: The estimated effect of feedback on the national rank in the twelfth grade by track/specialisation at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. Three tracks are available in all schools: Classics, Science and Information Technology. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 14: Treatment effects on the popularity position and rank of the program admitted conditional on prior performance



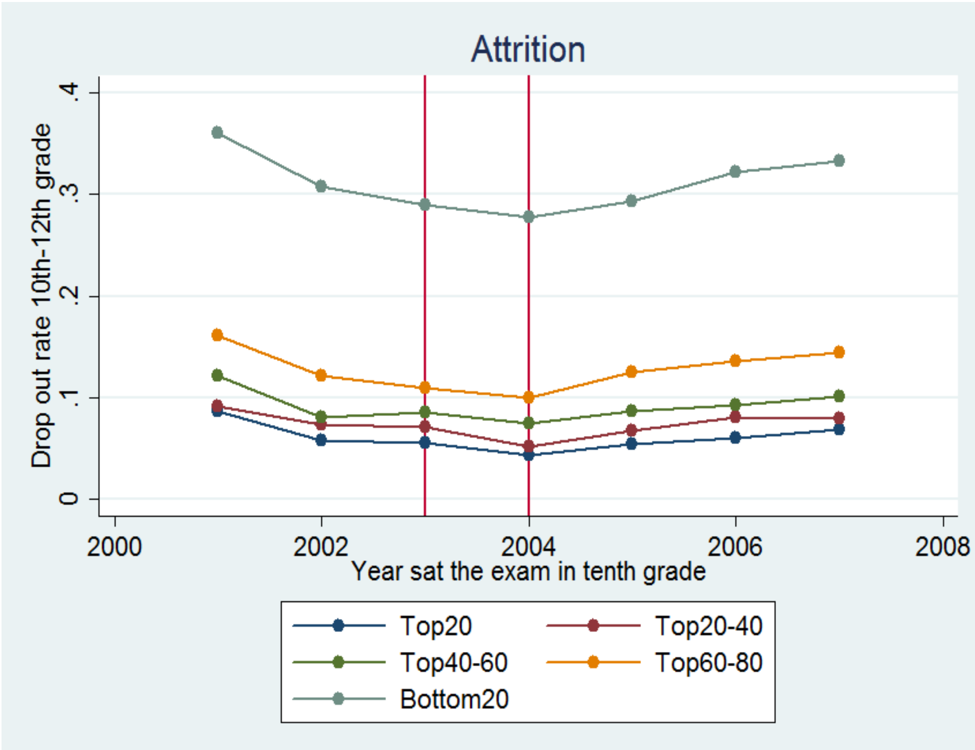
Note: The estimated effect of feedback on the popularity position (on the left) and rank(on the right) of the program admitted and the associated 95 % confidence interval. There are 672 programs in total. Popularity position and rank measured by the average University Department cut-off score over seven years. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 15: Treatment effects on the annual expected earnings conditional on prior performance



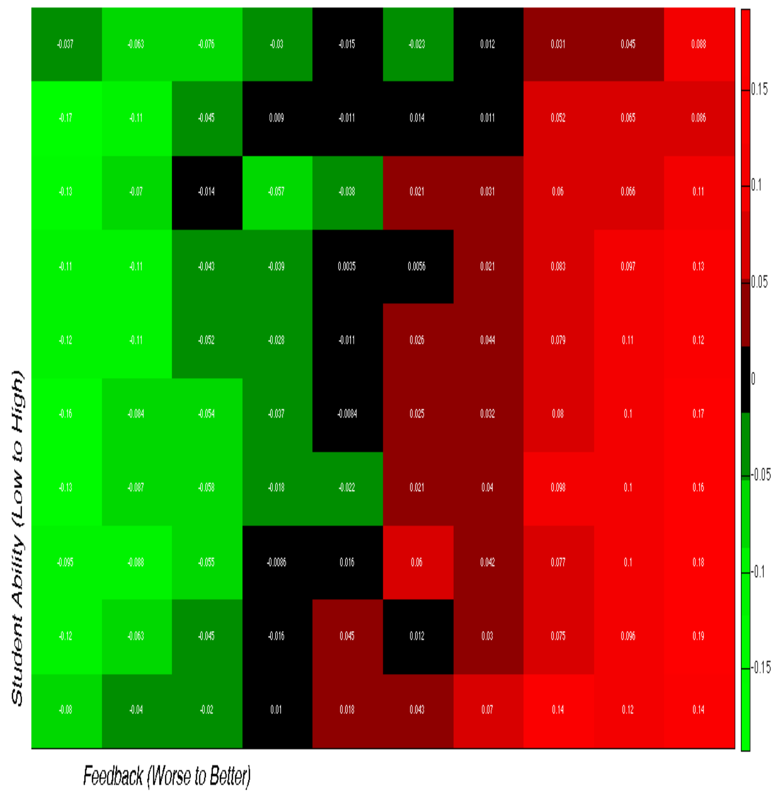
Note: The estimated effect of feedback on the expected annual wage at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The annual expected earnings are calculated based on the actual annual earnings of older graduates who studied the same college field. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

Figure 16: Drop out rates for each quintile of students' prior performance



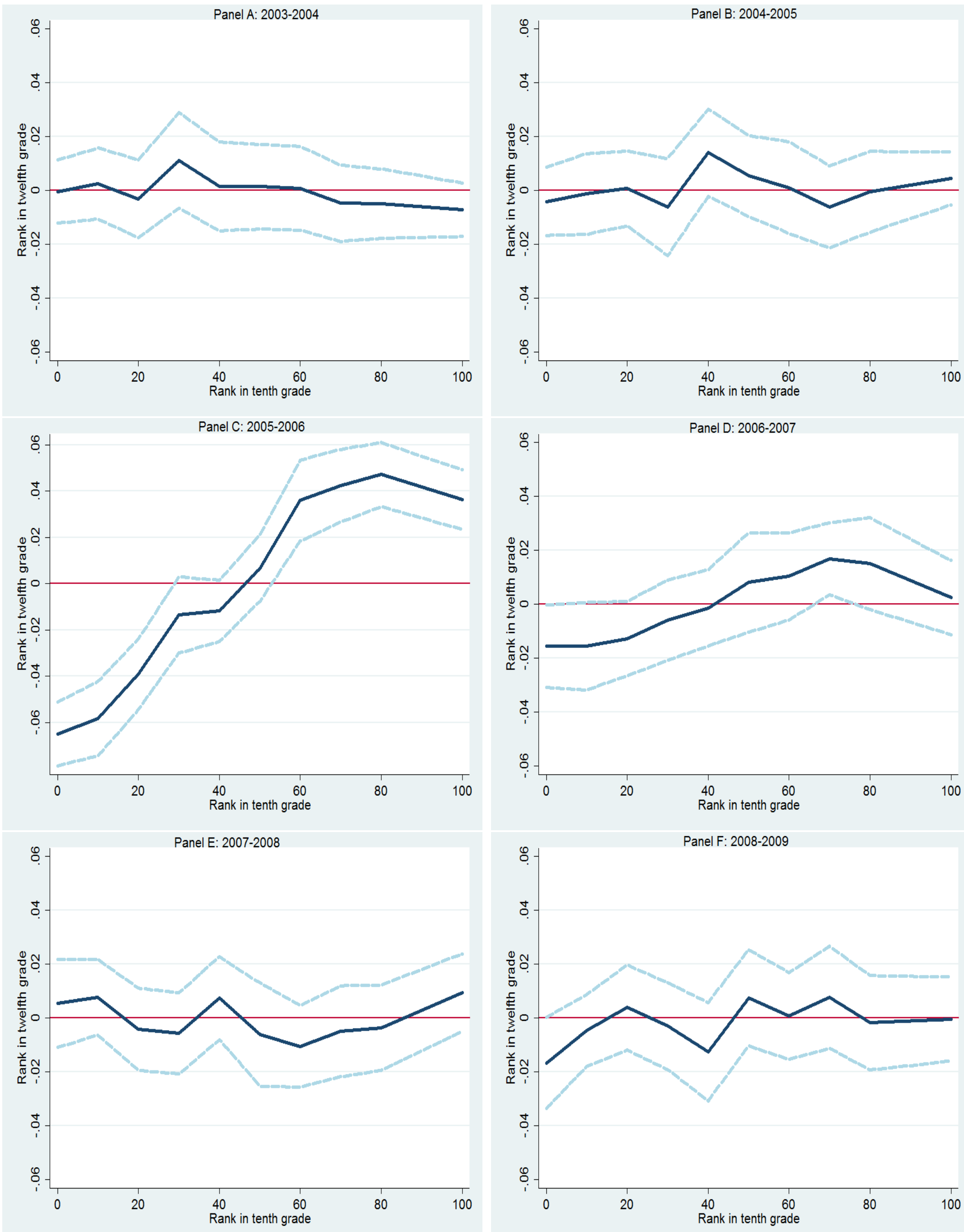
Note: Drop out rates between the tenth and the twelfth grade for each quintile of students' GPA performance in the tenth grade. Cohorts that are in tenth grade from 2001 to 2003 sit national exams in eleventh grade. Cohorts that are in the tenth grade after 2004 they do not sit national exams in the eleventh grade.

Figure 17: Positive and Negative Surprise



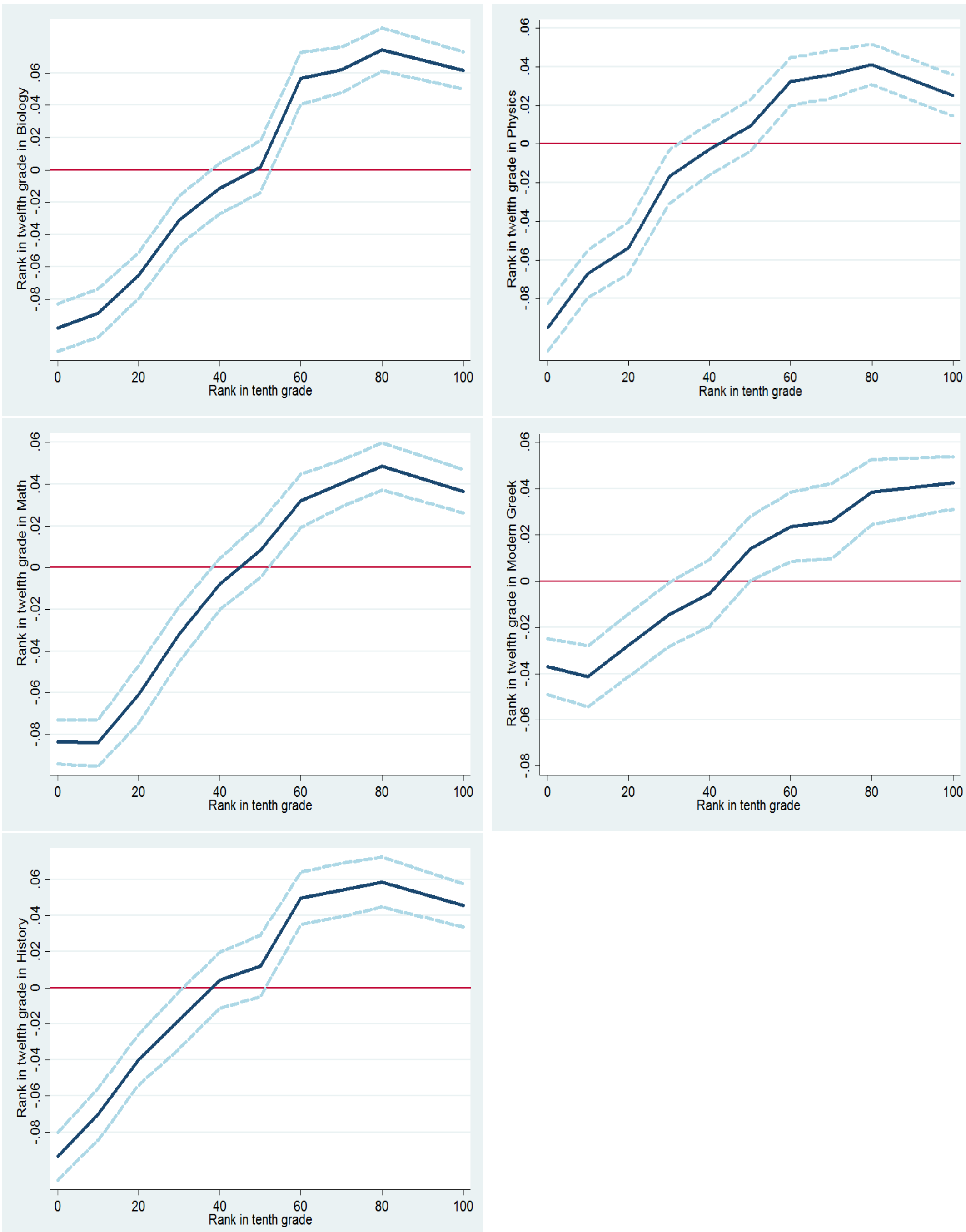
Note: Treatment effect for students with positive or negative surprise. Student performance (in deciles) in tenth grade on the vertical axis and student performance in eleventh grade (in deciles) on the horizontal axis.

Figure 18: Robustness checks



Note: Robustness checks: As if feedback was abolished in 2004 (Panel A), 2005 (Panel B), 2006 (Panel C), 2007 (Panel D), 2008 (Panel E) and 2009 (Panel F).

Figure 19: Feedback effects on twelfth grade rank nationwide for each subject separately conditional on prior performance



Note: The estimated effect of feedback on the twelfth garde rank nationwide at each decile of students' GPA performance in the tenth grade and the associated 95 % confidence interval. The regressions are conditional on the students' characteristics: gender, age, a dummy that takes the value of one if the student is early enrolled in school , school fixed effects and dummies for the track each student chooses in the twelfth grade. Standard errors are clustered at the school level.

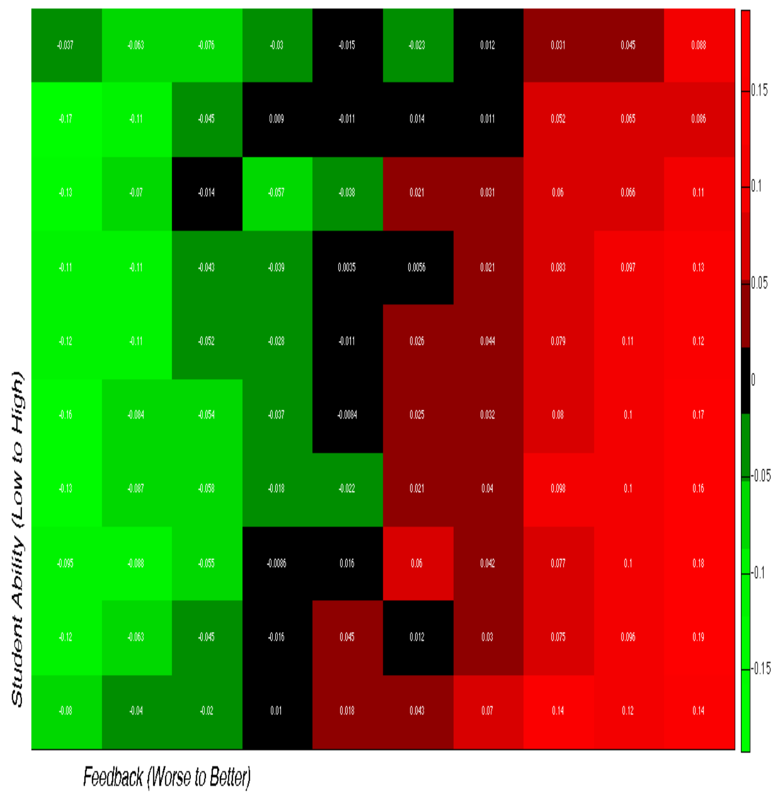


Figure 20: Treatment effect for students with positive or negative surprise. Student performance (in deciles) in tenth grade on the vertical axis and student performance in eleventh grade (in deciles) on the horizontal axis.

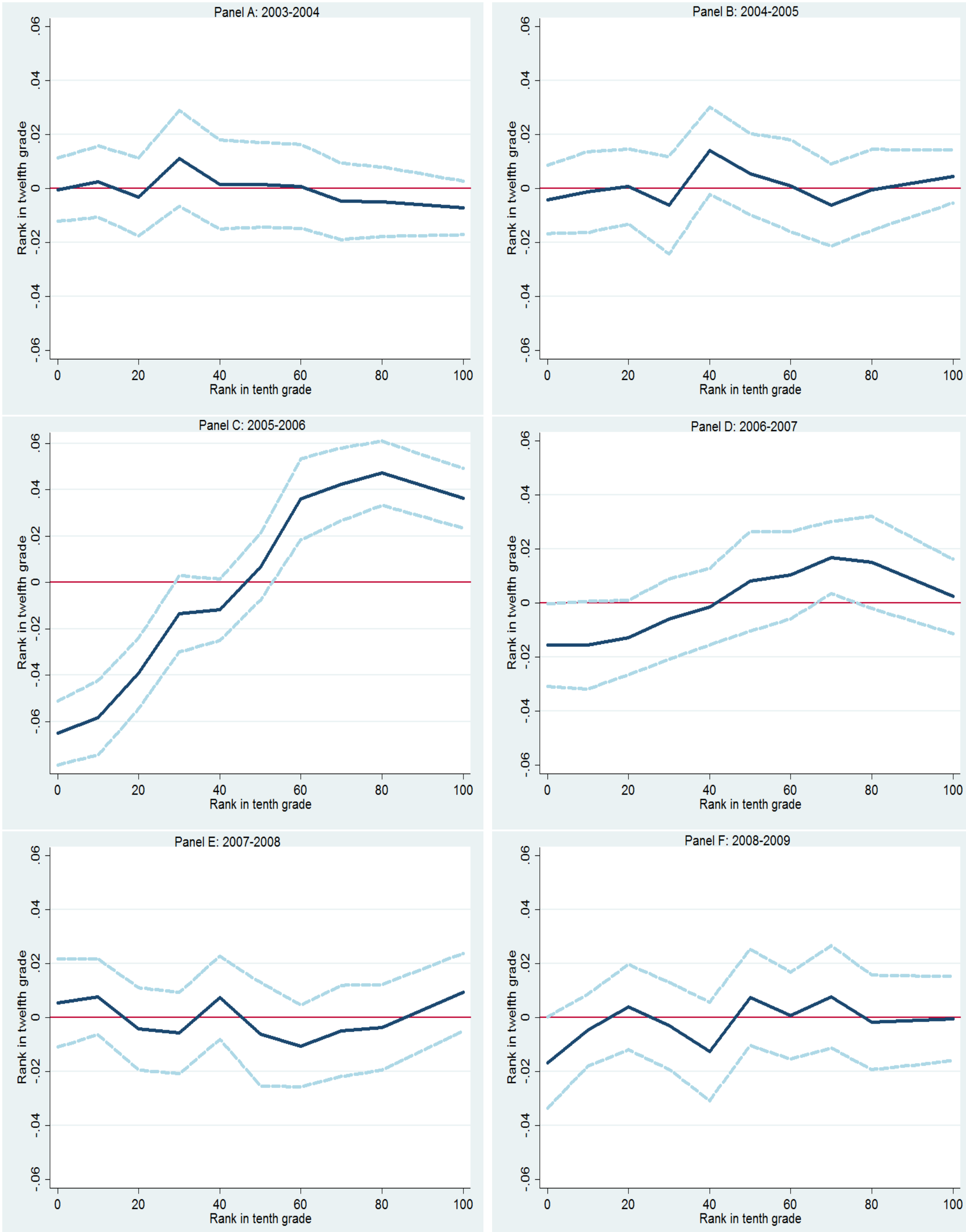


Figure 21: Robustness checks: As if feedback was abolished in 2004 (Panel A), 2005 (Panel B), 2006 (Panel C), 2007 (Panel D), 2008 (Panel E) and 2009 (Panel F).

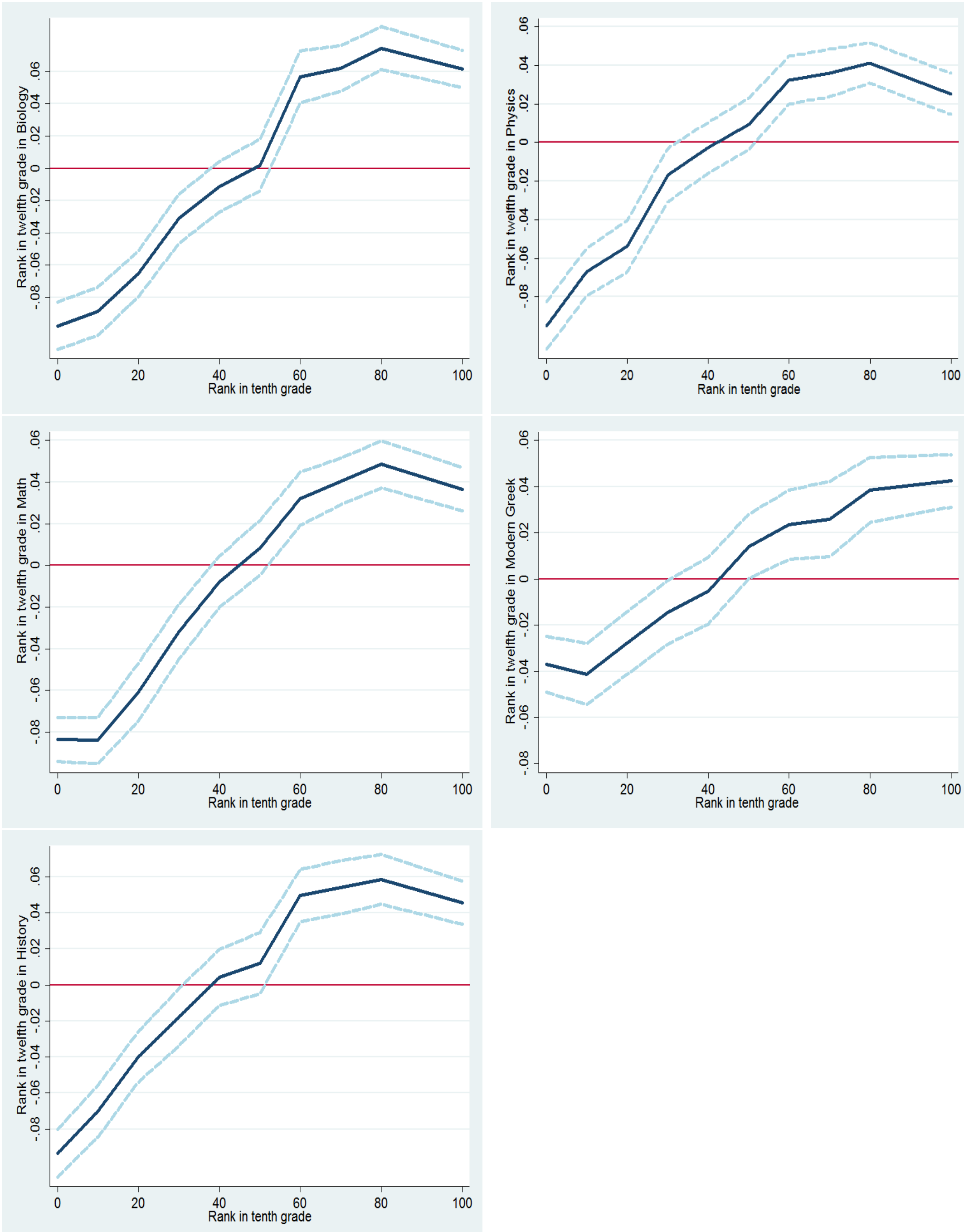


Figure 22: Feedback effects for each decile of prior ability rank by subject based on tenth grade GPA

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Student Characteristics in twelfth grade				
Age	17.875	0.466	17	27
Early enrollment	0.167	0.373	0	1
Female	0.566	0.496	0	1
School cohort size	78.518	31.17	10	170
School GPA	85.930	10.186	49.44	100
National exam grade	62.843	19.362	7.550	98.857
Cohort size	63,186	8,710	50,061	71,796
logIncome(in 2009 Euro)	9.999	0.270	9.473	11.105
Retake the national exam	0.115	0.319	0	1
Specialty Characteristics				
Specialty:Classics	0.359	0.48	0	1
Specialty:Exact Sciences	0.164	0.371	0	1
Specialty:Information Technology	0.477	0.499	0	1
School Characteristics				
Private School	0.039	0.193	0	1
Experimental School	0.061	0.24	0	1
Public School	0.9	0.3	0	1
Urban	0.973	0.161	0	1
University Admission				
Admitted	0.823	0.381	0	1
College district different from school district	0.677	0.468	0	1
Number of university departments	8.293	10.543	1	242
Rank of admitted college in preference list	24.699	21.618	1	254
Places in tertiary education	60,960	6,268	52,450	68,136

Note: 45.842 obs. 7 cohorts. The variable ”places in tertiary education” is calculated as the average across admitted students.

Table 2: Sample and Population

Variable	134schools Mean	1189schools Mean	Difference (b/s.e.)
Age	17.875	17.892	-0.017***
			(0.003)
Early enrollment	0.167	0.167	-0.0004
			(0.002)
Female	0.566	0.565	0.002
			(0.003)
School cohort size	78.518	75.358	3.160
			(0.197)
logIncome (in 2009Euro, annual)	9.999	9.938	0.060***
			(0.001)
Retake	0.115	0.112	0.003
			(0.002)
Specialty: Classics	0.359	0.366	-0.007
			(0.004)
Specialty: Exact Sciences	0.164	0.159	0.005
			(0.002)*
Specialty: Information Technology	0.477	0.475	0.002
			(0.003)
School and University Characteristics			
Private school	0.039	0.080	-0.041***
			(0.001)
Public schools	0.900	0.901	-0.001
			(0.002)
Experimental school	0.061	0.019	0.042***
			(0.001)
Urban	0.973	0.892	0.082***
			(0.002)
Admitted	0.823	0.803	0.020***
			(0.001)
Internal migration	0.677	0.800	-0.123***
			(0.002)
Rank of admitted college in preference list	8.293	8.584	-0.292***
			(0.065)
No of university departments in preference list	24.699	26.865	-2.166***
			(0.120)

Note: 45,842 obs. in sample and 431,469 obs. in population. There are in total 1,323 senior high schools in operation. Evening schools are excluded from the sample and the population

Table 3: Treatment and Control Group

Variable	Feedback Mean	No Feedback Mean	Difference (b/s.e.)
Student and Speciality Characteristics			
Age	17.835	17.909	0.074***
			(0.004)
Early enrollment	0.209	0.129	-0.080***
			(0.004)
Female	0.553	0.579	0.026***
			(0.005)
School cohort size	88.083	70.030	18.053***
			(0.288)
logIncome (in 2009Euro,annual)	9.988	10.005	0.017***
			(0.003)
Retake	0.104	0.124	0.020***
			(0.003)
Specialty: Classics	0.344	0.377	0.033***
			(0.004)
Specialty: Exact Sciences	0.176	0.154	-0.022***
			(0.004)
Specialty: Information Technology	0.480	0.469	-0.011**
			(0.005)
School and University Characteristics			
Private school	0.037	0.037	0.0003
			(0.002)
Public schools	0.905	0.897	-0.008
			(0.005)
Experimental school	0.058	0.066	0.007
			(0.005)
Urban	0.972	0.974	0.002
			(0.002)
Admitted	0.836	0.814	-0.022***
			(0.004)
Internal migration	0.475	0.635	-0.160***
			(0.002)
Rank of admitted college in preference list	9.657	7.068	-2.589***
			(0.115)
No of university departments in preference list	26.946	22.724	-2.589***
			(0.011)

Note: 21.965 obs. in treatment group and 23.781 obs. in control group. The feedback period is the pooled period from 2003 to 2005 while the non-feedback period consists of the pooled period from 2006 to 2009.

Table 4: Estimation results: Rank nationwide

Dependent Variable: Rank nationwide in incentivized subjects			
	Specifications		
Variable	(1)	(2)	(3)
Feedback*quintile5	0.042***	0.043***	0.045***
	(0.004)	(0.006)	(0.004)
Feedback*quintile4	0.036***	0.037***	0.040***
	(0.004)	(0.005)	(0.004)
Feedback*quintile2	-0.045***	-0.045***	-0.038***
	(0.004)	(0.004)	(0.005)
Feedback*quintile1	-0.088***	-0.088***	-0.079***
	(0.004)	(0.004)	(0.004)
Feedback	0.009***	0.009	-0.001
	(0.003)	(0.009)	(0.003)
quintile5	0.234***	0.235***	0.251***
	(0.003)	(0.004)	(0.004)
quintile4	0.094***	0.094***	0.102***
	(0.003)	(0.003)	(0.003)
quintile2	-0.081***	-0.083***	-0.093***
	(0.003)	(0.003)	(0.003)
quintile1	-0.176***	-0.177***	-0.192***
	(0.003)	(0.003)	(0.003)
Female	-0.008***	-0.008***	-0.011***
	(0.001)	(0.002)	(0.002)
Specialty: Science	0.055***	0.052***	0.048***
	(0.002)	(0.002)	(0.002)
Specialty: Classics	-0.022***	-0.021***	-0.021***
	(0.002)	(0.002)	(0.002)
Log Income	0.055***		
	(0.003)		
Experimental school	0.029***		
	(0.004)		
Private school	0.145***		
	(0.004)		
Urban	0.021***		
	(0.004)		
Year FE.	no	no	yes
School FE.	no	yes	yes
Observations	45,746	45,746	45,746
R squared	0.635	0.666	0.675
No of schools	134	134	134

Note: A constant is also included. Clusters at school level. *,**,*** denotes significance at the 10%,5% and 1% level respectively.

Table 5: Rank within the school in incentivized and non-incentivized subject

Dependent Variable: School Rank in incentivized and non-incentivized subjects				
	Incentivized subjects		**Non-Incentiv. subjects**	
Variable	(1)	(2)	(3)	(4)
Feedback*quintile5	0.045***	0.045***	0.005	0.005
	(0.004)	(0.004)	(0.006)	(0.006)
Feedback*quintile4	0.040***	0.040***	-0.005	-0.005
	(0.004)	(0.004)	(0.006)	(0.006)
Feedback*quintile2	-0.038***	-0.038***	-0.004	-0.003
	(0.004)	(0.005)	(0.006)	(0.006)
Feedback*quintile1	-0.079***	-0.079***	0.005	0.005
	(0.004)	(0.004)	(0.005)	(0.006)
Feedback	0.001	-0.001	0.003	0.001
	(0.003)	(0.003)	(0.004)	(0.004)
quintile5	0.251***	0.251***	0.256***	0.256***
	(0.004)	(0.004)	(0.005)	(0.004)
quintile4	0.102***	0.102***	0.103***	0.105***
	(0.003)	(0.003)	(0.003)	(0.004)
quintile2	-0.093***	-0.093***	-0.094***	-0.095***
	(0.003)	(0.003)	(0.004)	(0.005)
quintile1	-0.193***	-0.192***	-0.200***	-0.200***
	(0.003)	(0.003)	(0.004)	(0.006)
Female	-0.009***	-0.011***	0.054***	0.054***
	(0.001)	(0.002)	(0.003)	(0.002)
Specialty: Science	0.047***	0.048***	0.033***	0.034***
	(0.001)	(0.002)	(0.002)	(0.003)
Specialty: Classics	-0.019***	-0.021***	0.097***	0.097***
	(0.001)	(0.002)	(0.002)	(0.003)
Log Income	0.051***	0.049***	0.007	0.007
	(0.0004)	(0.001)	(0.004)	(0.004)
Experimental school	-0.041***	-0.038***	-0.003	-0.004
	(0.002)	(0.004)	(0.004)	(0.003)
Private school	-0.003	-0.004	0.030	0.032
	(0.003)	(0.004)	(0.016)	(0.018)
Urban	-0.017***	-0.016***	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Year FE.	no	yes	no	yes
Observations	45.746	45.746	45.746	45.746
R squared	0.674	0.675	0.542	0.543
No of schools	134	134	134	134

Note: Standard errors are clustered at the school level. A constant is also included. *,**,*** denotes significance at the 10%,5% and 1% level respectively.

Table 6: Estimation results : Differential Response by Gender

Dependent Variable: Rank in twelfth grade				
	Rank within the school		**Rank nationwide**	
Variable	(1)	(2)	(3)	(4)
Female*Feedback	-0.028	-0.028	-0.027	-0.027
	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Female	0.054	0.054	0.052	0.052
	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Feedback	0.009	0.002	0.009	0.008
	(0.003)***	(0.004)	(0.008)	(0.009)
Speciality in Science	0.198	0.199	0.198	0.196
	(0.004)***	(0.004)***	(0.004)***	(0.004)***
Speciality in Classics	-0.040	-0.039	-0.039	-0.040
	(0.003)***	(0.003)***	(0.005)***	(0.003)***
Income	-0.0001	-0.0001	0.0002	0.0002
	(0.0001)	(0.0001)	(0.0001)***	(0.0001)***
Private	-0.015	-0.015	0.134	0.134
	(0.008)*	(0.008)*	(0.016)***	(0.017)***
Experimental	-0.015	-0.015	0.017	0.017
	(0.006)**	(0.006)**	(0.018)	(0.018)
urban	-0.029	-0.029	0.007	0.007
	(0.007)***	(0.007)***	(0.015)	(0.015)
R^2	0.14	0.14	0.16	0.16
N	45,746	45,746	45,746	45,746
Year FE		✓		✓
No of schools	134	134	134	134

Note: Standard errors are clustered at the school level. A constant is also included. *,**,*** denotes significance at the 10%,5% and 1% level respectively. The rank in the twelfth grade here takes into account only the incentivized subjects. It is calculated within the school for columns (1) and (2) and across schools in columns (3) and (4)

Table 7: Capacity of schools

Variable	Mean	Std. Dev.	Min.	Max.
Schools with one class				
Public	0.899	0.302	0	1
Private	0.101	0.301	0	1
Experimental	0	0	0	0
Urban	0.378	0.485	0	1
Class size	18.130	5.717	10	29
No of schools	14			
No of students	522			
Schools with two classes				
Public	0.932	0.252	0	1
Private	0	0	0	0
Experimental	0.068	0.252	0	1
Urban	0.378	0.485	0	1
Class size	16.000	4.739	10	27
No of schools	38			
No of students	3,709			
Schools with three classes				
Public	0.941	0.235	0	1
Private	0.053	0.223	0	1
Experimental	0.006	0.077	0	1
Urban	0.986	0.115	0	1
Class size	18.211	4.998	10	32
No of schools	63			
No of students	9,959			
Schools with three classes				
Public	0.881	0.324	0	1
Private	0.035	0.184	0	1
Experimental	0.084	0.277	0	1
Urban	1	0	0	1
Class size	20.072	6.973	10	33
No of schools	74			
No of students	26,354			

Note: 111 senior high schools provided us with the eleventh and twelve grade classroom information. The number of classes in a school may not be stable across years. Some schools may expand and some others may shrink in some years.

Table 8: Loss of human capital in terms of labour force participants

Year	**Students Retaking**	**Potential Impact on Labour Market**
2003	7925	0.167%
2004	7223	0.150%
2005	6387	0.131%
2006	10421	0.213%
2007	6642	0.135%
2008	5730	0.116%
2009	4576	0.092%
2010	7680	0.153%

Table 9: Decision to Retake and Feedback

Dependent Variable: Repeat the national exams				
	LPM		**Probit**	**Logit**
Variable	(1)	(2)	(3)	(4)
Feedback* Misplacement	0.058	0.059	0.345	0.602
	(0.016)***	(0.016)***	(0.092)***	(0.181)***
Feedback	0.012	0.019	0.070	0.131
	(0.006)*	(0.007)**	(0.036)*	(0.074)*
Misplacement	-0.014	-0.015	-0.071	-0.099
	(0.014)	(0.015)	(0.077)	(0.142)
Age	-0.014	-0.019	-0.076	-0.157
	(0.003)***	(0.006)***	(0.039)*	(0.062)**
Early Enrolled	-0.005	-0.006	-0.011	-0.033
	(0.008)	(0.008)	(0.022)	(0.082)
Female	-0.007	-0.007	-0.044	-0.073
	(0.003)*	(0.004)*	(0.020)*	(0.038)*
Specialization in Classics	-0.020	-0.018	-0.113	-0.200
	(0.004)***	(0.007)*	(0.024)***	(0.046)***
Specialization in Science	0.013	0.016	0.090	0.169
	(0.005)**	(0.004)***	(0.026)***	(0.049)***
District Unemployment	0.005	0.002	0.025	0.046
	(0.002)**	(0.002)	(0.012)*	(0.019)**
If admitted in first place	-0.212	-0.218	-1.041	-1.964
	(0.008)***	(0.008)***	(0.035)***	(0.070)***
Internal Migration	0.064	0.072	0.445	0.889
	(0.005)***	(0.005)***	(0.037)***	(0.077)***
logIncome	-0.009			
	(0.011)			
Urban	0.024			
	(0.013)*			
Private	-0.056			
	(0.007)**			
Public	-0.039			
	(0.009)***			
R^2 or pseudo-R squared	0.05	0.06	0.07	0.07
Log likelihood			-13,432	-13,439
School FE		✓	✓	✓
Year FE	✓	✓	✓	✓
N	45,746	45,746	45,746	45,746

Note: A constant is also included. Standard errors are clustered at the school level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 10: Decision to Retake, Feedback and Misplacement

Dependent Variable: Repeat the national exams				
	LPM		Probit	Logit
Variable	(1)	(2)	(3)	(4)
Feedback	-0.031	-0.002	-0.007	-0.017
	(0.007)***	(0.008)	(0.047)	(0.090)
Feedback* Misplacement Quintile 5	0.045	0.040	0.219	0.412
	(0.010)***	(0.009)***	(0.050)***	(0.095)***
Feedback* Misplacement Quintile 4	0.023	0.023	0.120	0.231
	(0.010)**	(0.009)**	(0.049)**	(0.095)**
Feedback* Misplacement Quintile 2	0.004	0.007	0.049	0.103
	(0.010)	(0.011)	(0.054)	(0.101)
Feedback* Misplacement Quintile 1	-0.034	-0.031	-0.151	-0.274
	(0.010)***	(0.010)***	(0.052)***	(0.098)***
Misplacement Quintile 5	-0.017	-0.018	-0.103	-0.184
	(0.007)**	(0.007)**	(0.038)***	(0.073)**
Misplacement Quintile 4	-0.025	-0.025	-0.139	-0.262
	(0.007)***	(0.007)***	(0.038)***	(0.072)***
Misplacement Quintile 2	0.017	0.016	0.076	0.143
	(0.007)**	(0.008)**	(0.039)*	(0.073)**
Misplacement Quintile 1	0.030	0.031	0.148	0.273
	(0.007)***	(0.009)***	(0.043)***	(0.080)***
Female	-0.010	-0.010	-0.056	-0.105
	(0.004)***	(0.004)***	(0.020)***	(0.037)***
Age	0.002	-0.001	-0.001	-0.002
	(0.007)	(0.007)	(0.035)	(0.067)
Early Enrolled	0.011	0.009	0.047	0.087
	(0.008)	(0.008)	(0.041)	(0.078)
Unemployment	0.005	0.002	0.010	0.020
	(0.001)***	(0.002)	(0.011)	(0.021)
Internal migration	-0.024	-0.022	-0.109	-0.211
	(0.007)***	(0.008)***	(0.038)***	(0.075)***
Specialization in Science	-0.007	-0.004	-0.018	-0.036
	(0.005)	(0.005)	(0.025)	(0.048)
Specialization in Classics	-0.018	-0.017	-0.093	-0.175
	(0.004)***	(0.004)***	(0.024)***	(0.045)***
Private	-0.087			
	(0.011)***			
Public	-0.040			
	(0.009)***			
LogIncome	-0.033			
	(0.008)***			
Urban	0.006			
	(0.010)			
R^2 or pseudo-R squared	0.03	0.04	0.06	0.06
Log likelihood			-14,062	-14,063
School FE		✓	✓	✓
Year FE	✓	✓	✓	✓
N	45,746	45,746	45,746	45,746

Note: A constant is also included. Standard errors are clustered at the school level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 11: Estimation results : Drop out

Dependent Variable: Dummy for drop out		
	Specifications	
Variable	(1)	(2)
Feedback*quintile5	0.009	0.010
	(0.007)	(0.007)
Feedback*quintile4	0.007	0.007
	(0.007)	(0.007)
Feedback*quintile2	0.009	0.010
	(0.008)	(0.008)
Feedback*quintile1	0.013	0.014
	(0.015)	(0.016)
Feedback	0.017	0.041
	(0.019)	(0.033)
quintile5	0.000	-0.001
	(0.004)	(0.004)
quintile4	-0.006	-0.006
	(0.005)	(0.005)
quintile2	0.025***	0.025***
	(0.006)	(0.006)
quintile1	0.153***	0.153***
	(0.003)	(0.014)
Female	-0.011***	-0.011***
	(0.003)	(0.004)
Absences10	0.001***	0.002***
	(0.0001)	(0.0001)
Year FE.	no	yes
Observations	56.041	56.041
R squared	0.130	0.203
No of schools	134	134

Note: A constant is also included. Clusters at school level. *,**,*** denotes significance at the 10%,5% and 1% level respectively. Quintiles are constructed based on the school performance in tenth grade used.

Table 12: Estimation results: Different outcome variables

Dependent Variable: Rank in twelfth grade			
	Specifications		
Variable	(1)	(2)	(3)
Feedback*quintile5	0.026***	0.030***	0.050***
	(0.004)	(0.007)	(0.005)
Feedback*quintile4	0.022***	0.015***	0.032***
	(0.004)	(0.007)	(0.005)
Feedback*quintile2	-0.029***	-0.032***	-0.042***
	(0.004)	(0.007)	(0.005)
Feedback*quintile1	-0.052***	-0.045***	-0.066***
	(0.004)	(0.006)	(0.005)
Feedback	0.002	0.008	-0.0004
	(0.003)	(0.005)	(0.003)
quintile5	0.257***	0.247***	0.245***
	(0.003)	(0.005)	(0.003)
quintile4	0.109***	0.110***	0.107***
	(0.003)	(0.005)	(0.003)
quintile2	-0.097***	-0.100***	-0.091***
	(0.003)	(0.005)	(0.003)
quintile1	-0.207***	-0.231***	-0.210***
	(0.003)	(0.005)	(0.003)
Female	-0.019***	0.030***	-0.014 ***
	(0.001)	(0.002)	(0.001)
Early Enrollment	0.010***	0.009***	0.011 ***
	(0.002)	(0.002)	(0.002)
Specialty: Science	0.006***	0.019*** *	0.023***
	(0.002)	(0.004)	(0.002)
Specialty: Classics	0.010***	0.098***	-0.059
	(0.002)	(0.003)	(0.002)
Observations	45,746	45,746	45,746
R squared	0.661	0.674	0.625
No of schools	134	134	134

Note: A constant is also included. The outcome in the first column is the rank calculated based on the five core subjects and the four Track subjects. The outcome in the second column is the rank in Modern Greek. The outcome variable in the third column is calculated based on five subjects in the feedback regime and two subjects in the non-feedback regime. Standard errors clustered at the school level. Year fixed effects included. Clusters at school level. *,**,*** denotes significance at the 10%,5% and 1% level respectively.

Table 13: Descriptive Evidence of Social Mobility

Quintiles of program' popularity	Quintiles of Neighborhood Income									
	Quintile1		**Quintile2**		**Quintile3**		**Quintile4**		**Quintile5**	
	Feed.	No Feed.	Feed.	No Feed.	Feed.	No Feed.	Feed.	No Feed.	Feed.	No Feed.
Quintile1	4.6	3.4	4.2	3.2	4.5	2.9	2.7	2.1	3.1	2.7
	+1.2		+1		+1.6		+0.6		+0.4	
Quintile2	3.9	3.5	3.5	3.1	4.1	3.5	2.5	2.7	3.1	3.2
	+0.4		+0.5		+0.6		-0.2		-0.3	
Quintile3	3.6	3.4	3.2	3.2	3.8	3.5	2.7	3	3.2	3.4
	+0.2		0		+0.3		-0.3		-0.3	
Quintile4	3.1	3.0	2.8	3.2	3.5	4.3	2.5	3.2	3.4	3.8
	+0.1		-0.4		-0.8		-0.7		-0.4	
Quintile5	2.9	2.6	2.6	3.0	3.9	4.1	2.8	3.2	3.6	4.2
	+0.3		-0.42		-0.2		-0.4		-0.6	
Total	18.1	15.9	16.3	15.7	19.8	18.3	13.2	14.2	16.3	17.4
	2.2		**+0.6**		**+1.5**		**-1**		**-1.1**	

Note: The variable "Quintile 1" represents the bottom quintile of program's popularity and neighborhood income. The "Quintile 5" denotes the top quintile of the program's popularity and neighborhood income. For each quintile of neighborhood income two percentages are reported: the first one corresponds to the feedback period and the second one to the non-feedback period. The differences between the percentage in the feedback period and the non-feedback period for each quintile of program's popularity are also reported.