

IZA DP No. 4051

Individual Teacher Incentives, Student Achievement and Grade Inflation

Pedro S. Martins

March 2009

Individual Teacher Incentives, Student Achievement and Grade Inflation

Pedro S. Martins

*Queen Mary, University of London,
CEG-IST and IZA*

Discussion Paper No. 4051
March 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

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

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

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

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

ABSTRACT

Individual Teacher Incentives, Student Achievement and Grade Inflation^{*}

How do teacher incentives affect student achievement? We contribute to this question by examining the effects of the recent introduction of teacher performance-related pay and tournaments in Portugal's public schools. Specifically, we draw on matched student-school panel data covering the population of secondary school national exams over seven years. We then conduct a difference-in-differences analysis based on two complementary control groups: public schools in two autonomous regions that were exposed to lighter versions of the reform than in the rest of the country; and private schools, which are also subject to the same national exams but whose teachers were not affected by the reform. Our results consistently indicate that the increased focus on individual teacher performance caused a significant decline in student achievement, particularly in terms of national exams. The triple-difference results also document a significant increase in grade inflation.

JEL Classification: I21, M52, I28

Keywords: performance-related pay, public sector, matched school-student data

Corresponding author:

Pedro S. Martins
School of Business and Management
Queen Mary, University of London
Mile End Road
London E1 4NS
United Kingdom
E-mail: p.martins@qmul.ac.uk

^{*} I thank Pedro Pita Barros, Alvaro Novo, Pedro Portugal and Helena Martins for helpful discussions. Any errors are of my responsibility only.

1 Introduction

Recent research has emphasized the importance of teachers in terms of students' achievement (Rivkin et al. 2005, Rockoff 2004, Martins & Walker 2006). However, this line of enquiry has also shown that it is particularly difficult to explain differences in teacher quality, at least when using teacher characteristics available for research analysis. These results have generated increased interest in the study of mechanisms that may raise achievement through teachers, including increasing teacher training, improving teachers' pay or strengthening teacher incentives (Lazear 2003). This paper addresses the latter policy option.

In fact, teacher incentives, either individual or collective, may improve student achievement if they succeed in aligning the public or even social goals with the goals of the teacher. In this case, a combination of incentive and composition effects will increase student performance (Lazear 2000, 2003). However, an approach in which reward is based on outputs can also be fraught with difficulties, which may explain the popularity of simpler input-based rewards (Kane & Staiger 2002). For instance, setting specific measurable outputs may lead to potentially dysfunctional behaviour such as teaching to test. Moreover, while individual incentives may disrupt collaborative work (Martins 2008), collective incentives may also generate free riding and, in the end, little incentive effect. Finally, the extra risk in pay in output-based compensation will need to be compensated by higher wages, especially when the output is determined by many other variables than the agent's effort, as in the case of education production.

Although the theoretical analysis on the effects of teacher incentives is ambiguous, the empirical literature in this area is particularly thin. We know of only three studies that explore randomized or quasi-natural experiments to address the causal relationship between teacher incentives and student achievement - Lavy (2002), Glewwe et al. (2003) and Lavy (forthcoming). Furthermore, the results of these three papers are not entirely consistent: the two studies based on Israel (Lavy 2002, forthcoming) are supportive of the potential of collective and individual teacher incentives, while the randomized experiment based on Kenya (Glewwe et al. 2003) is not.¹

In this paper, we add to this small literature by examining the effects of the recent intro-

¹See also Eberts et al. (2002) for a case study of a US county, Figlio & Kenny (2007) for a study of the US based on cross-sectional data and Atkinson et al. (forthcoming) for a recent analysis of the introduction of performance-related pay for teachers in England.

duction of individual teacher incentives in Portugal's public-sector schools. The main aspects of this reform, described in more detail in Section 2, are the breaking up of the until then single pay scale for teachers into two; and the performance-related criteria set for progression between the two pay scales. This established a clear contrast in the incentives faced by public-school teachers as until then salaries essentially only varied with tenure. Moreover, the criterion for teacher progression which received by far most attention, in the media and amongst teachers, was the (internal and external) results of the students taught by each teacher. Another much-debated criterion was the feedback from the students' parents about the teacher: if positive, the chances of progression would increase. In addition, all progressions between the two pay scales were also conditional on not exceeding a given number of vacancies per school, determined for all schools every two years by the Education Department. Finally, those teachers that did particularly well along the progression criteria and also met centrally-determined quotas would receive a one-off financial award. These and other aspects of the reform are broadly consistent with the characteristics of tournaments, as in Lazear & Rosen (1981).

We study the effects of these reforms in terms of students' school-level and national-exam results. Specifically, we draw on freely-available matched student-school data covering the population of secondary school students that sat national exams from 2002 to 2008 in Portugal. We then conduct a difference-in-differences analysis (Meyer 1995) based on two complementary control groups. In the first control group, we consider public schools in the Azores and Madeira. These are two autonomous regions of Portugal that were exposed to lighter versions of the reform than the rest of the country, as their pay scale remained unchanged and progression was less restricted. In the second control group, we consider private schools. These schools are again subject to the same national exams as the treatment group but whose teachers were not affected by the reform, as pay and other incentives are set freely by each private school.

Our paper contributes to the literature on the effects of teacher incentives in different ways. First, this is the first paper in this literature that examines a reform that was applied across a country (rather than in a pilot study) and that conducts the analysis drawing on representative (population) data. In this way, we are able to address most issues of external validity that typically receive little attention in the literature. More generally, our results

contrast with the approach common in the empirical literature on incentives which typically faces severe data constraints and therefore tends to be based on case-studies of individual organisations (Lazear 2000, Bandiera et al. 2005).

Our population data also contrast with the case of the three papers mentioned above that draw on randomized or quasi-experimental studies (Lavy 2002, Glewwe et al. 2003, Lavy forthcoming). In the words of Meyer (1995), "[a] weakness of natural experiments is that their results may not be generalizable beyond the group of individuals or firms or the setting used in the study." (page 159). Indeed, these papers do not claim to present representative evidence of the countries examined there as their research design generates ATT (average treatment on the treated) effects instead. Moreover, the fact that we are using official data throughout the period analysed improves the quality of our findings in terms of measurement error bias.

A second important aspect of our study is that as the assignment of schools to treatment is not random in our analysis, we conduct considerable robustness analysis, on top of our consideration of two complementary control groups. For instance, our analysis of a long period of up to five years before the reform allows us to assess the common trends hypothesis in some detail. Furthermore, we also consider a number of different specifications that control for different sets of variables, including school and school-exam fixed effects, and we also examine different data subsets and different aggregation levels.

Finally, we pay particular attention to the potential for grade inflation. This may result from the fact that the progression criteria are affected by student results that are determined at least in part by teachers themselves (see Jacob & Levitt (2003) for an extreme example of grade inflation). As we have individual student data on both school- and national-level results, we can examine grade inflation from a triple-differences perspective by considering, for each student-module pair, the difference between the internal and external mark. These results complement our evidence on the effects of the reform separately in terms of internal and external grades.

Overall, our results consistently indicate that the increased focus on individual teacher performance caused a sizable and statistically significant decline in student achievement. This decline in achievement is also much more pronounced in the case of national exams with an effect of up to 40% of a standard deviation. As in the different effects in terms of internal

and external results, our triple-difference evidence also documents a significant increase of grade inflation. In addition, in support of a causal interpretation of our results, we also find that in almost all specifications and dependent variables there are no significant differences between the treatment and control groups over time before the introduction of merit-pay. Finally, the inclusion of different control variables or the consideration of different subsets of the data makes only very minor differences to the size of our estimates, as would be the case if assignment to treatment were random.

The structure of the paper is as follows: Section 2 describes the main characteristics of the education reform studied in the paper and discusses some of its theoretical implications. Section 3 presents the data used in the paper, a matched school-student panel data set; Section 4 describes the main results, while Section 5 presents the robustness analysis. Finally, Section 6 concludes.

2 Background

A new government that came into office in 2005 decided to respond to the evidence of relatively poor performance levels in the Portuguese education system (OECD 2001). Indeed, when measured in terms of international comparisons such as the OECD's PISA tests, students in Portugal do not fare well. This is particularly true when taking into account the relatively high expenditure levels in education in the country, of which relatively high average teacher salaries are an important component.

The main aspect of the education reform was the breaking up of the until then single pay scale for teachers into two separate scales. This and other aspects of the reform became law in January 2007, after having been subject to public discussion for several months and approved by the government in November 2006. The breakup of the pay scale marked an important contrast with the period before the 2006/07 school year as teachers were no longer ensured of virtually automatic, tenure-related progression from the bottom to the top of the scale during their careers. In particular, the gap between the last point in the lower scale and the first point of the higher scale was potentially particularly large, at about 25%, from about €2,000 to about €2,500 per month (gross). On the other hand, those teachers in the higher pay scale had a different job title and were supposed to play a special role in management and pedagogical tasks in their schools.

Another key aspect of the reform is that the new policy conditioned progression from the lower to the upper pay scale on a number of aspects of individual teacher performance. These broadly measurable criteria were virtually inexistent until them. One such criterion for teacher progression, which received by far most media attention, was the results of the students taught by each teacher. Another criterion that also received considerable attention was the feedback from the students' parents about the teacher. The remaining criteria included the teacher's attendance record, attendance at training sessions, management and pedagogical duties, and involvement in research projects.

According to the law, these criteria for progression were to be assessed at each school, by those teachers in the higher pay scale. Moreover, detailed assessment sheets were made available by the Education Department and were supposed to be used in the process of gathering information regarding the above-mentioned criteria.² However, if the teacher did well along these criteria, progression between the two pay scales was also conditional on a given number of vacancies per school, determined centrally every two years by the Education Department as a function of the number of students in the school.

Finally, those teachers that did particularly well along the progression criteria and met centrally-determined quotas would also receive a one-off financial award. The specific amount of this award was not made public when the first law was released (January 2007) but recent reports as of January 2009 indicate that the annual award can amount to around one monthly salary.

On the political economy side, the reform generated considerable heated debate and opposition from teachers and their unions, including two national strikes. The extensive bargaining between government and unions that ensued created some confusion about the ultimate format of the incentive system, which is not yet absolutely clear at the time of writing.

However, several reports from different sources indicate that teachers acknowledged the change in incentives and adapted their behaviour in some way. Additionally, teachers without tenure and teachers with intermediate levels of tenure have already been assessed for the purpose of progression in the terms of new law. Their assessment includes the period covered by our data (see Section 3). New assessment rounds (which are to take place every two years)

²Teachers also complained that the administrative burden involved in the teacher assessment process was considerable. The time spent handling the formal aspects of applying for progression may also have reduced the effort that teachers put into teaching activities.

are being prepared at the moment and the period of time that will be taken into account for the purpose of that assessment again includes our data.

From the above, we conclude that the reform under study here involves a stark contrast in terms of teacher assessment and incentives. In particular, the new assessment framework introduces several aspects which can be characterised not only as performance-related pay but also more specifically as tournaments (Lazear & Rosen 1981). In this context, and turning to a theoretical discussion of the predicted effects of this reform, there are different arguments to take into account. On the one hand, the greater weight placed on performance indicators will presumably induce teachers to focus their effort on those criteria highlighted in the law. This is expected to increase student achievement, which is proxied by national exams and school-level results in our study.

On the other hand, tournaments are known to be potentially disruptive in terms of the collaborative work amongst agents involved in a competition (see Martins (2008) and the references therein). Moreover, collaborative work may be particularly important in the education context. Fairness concerns may come to the fore and undermine teacher morale (Fehr & Schmidt 1999). Second, there will be extra risk in pay which presumably needs to be compensated by higher wages. However, this was not the case, as pay scales were not improved. Adjustment may thus have taken place via worker turnover and worker composition, again with a potentially detrimental effect upon student achievement.

Finally, setting broadly measurable outputs may lead to dysfunctional behaviour such as teaching to test or engaging in grade inflation. In the specific case of the secondary schools considered here, it is worthwhile to point out that the final grade of each student results from a weighted average of the internal and external marks, the weights being 70% and 30%, respectively. Because of these weights and also because teachers have direct control of internal marks, evidence of grade inflation, if any, is more likely to be observed in terms of internal marks.

One important final aspect to refer in terms of the education reform studied here is that it only applied to a smaller extent in the cases of the two autonomous regions in Portugal, the Azores and Madeira. Indeed, these two regions have special legislative powers in some domains, including education, and they have decided not to follow the education reform in

the mainland.³ Specifically, the two regions also introduced greater emphasis on assessment of a teacher’s performance, under broadly the same criteria as in the mainland, although less so in the case of Madeira. However, one important difference that applied to both Azores and Madeira is that they did not break up their pay scale. These differences in the intensity of the treatment are exploited in our empirical analysis.

3 Data

Our data cover the population of high-school national exams in Portugal over seven school years, from 2001/02 to 2007/08. The data are made available by the National Exams Committee (JNE, *Jurí Nacional de Exames*), an agency of the Ministry of Education which is responsible for all matters regarding national exams carried out in the country. These national exams, studied in this paper, are required for the award of the high-school diploma and also for university entry (European Commission 2007).

The data also include information about the internal grades obtained by students in each module (a specific discipline of study, such as Portuguese or Maths) from their schools, which are based on test results and other criteria adopted by each school and teacher. There is also information about the students’ final result in each module, after taking into account each student’s internal and national-exam grades. Internal grades are censored below 10, the pass level, in a scale of 0 to 20, in which case the student cannot sit the national exam, except in special circumstances. All data used in the paper are freely available from the JNE website.⁴

The data do not include any individual student identifier; instead, each observation concerns a student-module-school-year combination. Typically, there will be several observations for each student but it will not be possible to match them. However, all schools and all modules are identified by name and through unique time-invariant codes. Importantly, there is also information on the school’s location at the *concelho* level, resulting in about 400 different geographical areas, and the school’s public or private status. Each student-module-school-year combination is also characterised in terms of several variables, namely if the exam is a

³The relevant legal documents are Law (*Decreto-Leis*) 17/2007, of January 19th and 200/2007, of May 22nd, and Regional Laws (*Decretos Legislativos Regionais*) 28/2006/A, of August 8th; 21/2007/A, of August 30th; and 6/2008/M, of February 25th.

⁴Link: <http://sitio.dgidc.min-edu.pt/JNE/Paginas/estatistica.aspx> (in Portuguese). The data were originally released openly so that the media could compile school rankings. As far as we know, the data have never been used for other purposes until now.

resit (either because the student failed before or because the student wants to improve their grade), if the student is applying for admission to university, and if the student is sitting the exam but is not enrolled in the school. The data also include the student's gender and age, but only for the last three years (2005/06-2007/08); and the student's school year when taking the module (typically 12th, which is the last in secondary education in Portugal, but also the 11th, as some modules are subject to national exams at that stage).

We create our main sample of analysis by drawing on all student-exam pairs that meet the following five conditions: a first sit in the first call of a student that is applying to university and is also enrolled in the module of the exam in the school where they are sitting the exam; only schools that are present in the data in all seven years. These criteria are similar to those adopted by most media when compiling school rankings. Our criteria are also imposed in order to ensure that the effect in terms of internal and external grades are based on the same sample and thus are comparable. The resulting 1,411,246 observations are distributed across 563 schools, of which 469 are public schools and subject to the reforms described in the previous section.⁵

Table 1 presents some descriptive statistics based on the student-exam level data. We find that the average internal mark is 13.3 while the exam mark is 10.6.⁶ These statistics are evidence of different standards between school and national exams and, perhaps, also of grade inflation in schools. Indeed, the mean of the difference between the internal and the external marks, 2.6, is not only positive but also considerably large when compared to the average mark of the internal exam. We also find that about 11% of the exams pertain to private schools and about 5% are from schools in the Azores and Madeira regions. Moreover, there is a downward trend in the number of exams in the period covered, which is consistent with the declining number of students enrolled in secondary school as indicated by national statistics. The exception to the trend is 2006, when new exams were introduced while some of the older exams were still sat by students.

Table 2 presents additional descriptive statistics, but referring to the school-level data. Of

⁵The original size of the data is 3,092,507. 28.68% of these observations refer to second calls; 31.22% are not enrolled in the school; 7.73% are not applying for university admission; and 26.98% are resitting the exam. Of course, these exclusion categories overlap for many observations. Only one school is not present in the data in all years, a result that highlights the stability of the education sector in Portugal. No school switches between public and private status. Moreover, extensive robustness analysis was conducted and the results presented below in Section 4 are not sensitive to different sample definitions as discussed in Section 5.

⁶The words *grade* and *mark* are used interchangeably here. They both refer to the result of a student, either at the school level or in national exams, and are expressed in the 0-20 scale.

interest for the interpretation of the magnitude of our main results, the standard deviations of the logarithms of the internal and external exams are 0.05 and 0.15, respectively. Moreover, the standard deviation of the difference between the internal and external result is 1.028. We also find that, on average, there are 417 exams per year per school.

4 Results

As mentioned before, we estimate the effects of the introduction of performance-related pay from difference-in-differences models of student grade equations. Our identification assumption is that there is no effect specific to public schools (in the continent) from the 2006/07 school year other than the education reform. Specifically, in the case of our first control group, schools located in Azores and Madeira, we estimate specifications as follows:

$$y_{ijt} = \beta_0 + \beta_1 \text{Continent}_j + \beta_2 \text{After}_t + \beta_3 \text{Continent}_j \times \text{After}_t + u_{ijt}. \quad (1)$$

In most cases, y_{ijt} denotes the logarithm of the grade (internal or external) of the student-exam pair i in school j in year t . In two alternative specifications, the dependent variable is instead a measure of grade inflation, namely the difference between the internal and the external grade of the same student-exam pair (a triple differences specification) or a dummy variable taking value one if the internal grade exceeds the external grade.

We do not consider the logarithm of the differences between the internal and external grades in order to keep in our analysis those cases of ‘grade deflation’, when the external grade is less than the internal grade. Our analysis of different nonlinear functions of the dependent variable also serves a useful robustness purpose. Indeed, if there are other relevant interaction effects that break down the identification assumption, our qualitative results are less likely to hold across different dependent variables. More important, the triple-difference specification is based on a weaker identifying assumption. It simply requires that there is no contemporaneous shock that affects the relative outcomes of the treatment group in the same years as the education reform.

In all cases, Continent_j is a dummy variable with value one if school j is located in mainland, continental Portugal (henceforth referred to as *continent*): this variable will pick up permanent differences in terms of the dependent variable between those schools located in

the continent and those located in Azores or Madeira. $After_t$ is another dummy variable, with value one if year t is 2007 (i.e. school year 2006-2007) or later, the period when the reforms were more intense, as discussed in Section 2: this variable will pick up across-the-board differences between the period before the intervention and the period after the intervention (the introduction of performance-related pay).

Finally, $Continent_j \times After_t$ is the product of the two previous dummy variable and its parameter, β_3 , is the object of interest in this paper. Its estimate will pick up the effect of the education reforms upon student achievement or grade inflation, i.e. any additional difference between the two types of schools that emerges after the intervention.

From the benchmark specification in equation 1, we consider three extended versions with different additional controls. The first version includes controls for school size (the total number of exams sat in each year) and regions (20 fixed effects). The second specification includes school size and school fixed effects (as the latter are perfectly collinear with the regions fixed effects, these region fixed effects are dropped from the analysis). Finally, the third additional specification includes school size and school-exam fixed effects.⁷ Importantly, all models are estimated with robust standard errors, allowing for clustering at the school level.

The first set of results, based on internal grades, are presented in Table 3. These results draw on the student-level data described in Table 1, except that private schools are dropped. Across all four specifications, we find negative point estimates, indicating that the levels of achievement of public schools in the continent fall with respect to public schools in Azores and Madeira. The magnitude of the estimates is very similar across the specifications, ranging from -.008 to -.01, or about one fifth of a standard deviation of the dependent variable across all schools. On the other hand, the statistical significance of the coefficients is poor, as it does not meet the 5% significance level.

However, when we turn to Table 4, which presents similar specifications but considering external grades instead, we find much higher effects, ranging from -.04 to -.054. These correspond to approximately one third of a standard deviation of the log of average grades across schools. In addition, all estimates here are also significant, even at the .1% level. The

⁷Because the structure of exams changes over the period, we focus on a subset of four of the most important topics covered in these tests: Portuguese, Maths, Biology, and Physics/Chemistry. Each subset includes all exams that include the topic above in their exam title.

comparison of the two sets of results (internal and external) therefore indicates that while national-exam results of public schools in the continent fall significantly and by a meaningful size with respect to the same change for public schools in the Azores and Madeira, the equivalent effect for internal marks is much smaller and much less precise.⁸

The contrast between the internal and external results suggests that grade inflation in the public schools in the continent is a presumably unintended effect of the reform. Indeed, our triple-difference estimates - see Table 5 - indicate precisely that. We find that the average gap between internal and external marks increases by .217 to .246 (or about one quarter of a standard deviation of average grade inflation) in public schools in the continent with respect to their counterparts in Azores and Madeira. In all cases the coefficients are significant, at least at the 5% level. Similarly, we also find evidence that the probability of grade inflation increases in the continent with respect to Azores and Madeira - see Table 6.

The fact that the latter estimates are relatively small (between .01 and .022, although all significant at the 5% level except in one case) indicates that the effects in terms of grade inflation were felt mostly along the intensive margin, i.e. increased inflation in marks that previously tended to exhibit inflation, rather than the extensive margin. This is related to the limited room for expansion along the extensive margin, as only 14% of the marks exhibited ‘deflation’ and only 23% of the marks are the same at the internal and external level (here defined as an absolute difference of less than .5).

We now turn to our second and complementary control group: private schools. Although it is well known that private schools tend to exhibit better results in terms of the academic achievement of their students, our difference-in-differences approach will control at least for permanent differences between them and public schools. (See Neal (1997) for a study of student achievement in Catholic private schools, which correspond to a large share of the private schools in our data.)

Similarly to the case of the first control group (equation 1), here we estimate the following difference-in-differences specification:

$$y_{ijt} = \beta'_0 + \beta'_1 Public_j + \beta'_2 After_t + \beta'_3 Public_j \times After_t + u_{ijt}, \quad (2)$$

⁸We also find, both in Table 3 and in Table 4, that the *After* coefficients are always significantly positive, suggesting a trend towards higher marks.

All variables take the same interpretation as before; and $Public_j$ is a dummy variable with value one if school j is a State school. β'_3 is now the parameter of interest. All models in this section are estimated with the full set of student-level data described in Table 1, except that schools located in the Azores and Madeira islands (public or private) are dropped.

Table 7 presents the results for the internal grades. As in the case in which the public schools in Azores and Madeira served as the control group, we find again evidence that the introduction of the individual teacher incentives had a detrimental effect on student achievement. However, in the present case, the coefficients range from -.017 to -.022, or about twice as much as in the equivalent specification under the first control group. Moreover, the coefficients are now always statistically significant, even at the .1% level. The increase in the magnitude of the effects, to around 40% of a standard deviation, is also consistent with the increased intensity of the treatment in public schools in the continent when compared to public schools in Azores and Madeira, as discussed in Section 2.

These findings are also consistent with the external grade results - see Table 8. There we find that achievement in public schools, when compared to private schools, falls by between -.046 to -.065, and is again always significant at the .1% level. The magnitude of these effects corresponds to about one third of a standard deviation of external results.

Finally, the grade inflation result, obtained from the stronger effects in external grades when compared to internal grades is once again corroborated from the triple-difference results. In Table 9, we find that grade inflation increases by between .193 to .27 across the four specifications considered there, all of which are significant at the 1% level. Similar results to those of the first control group are found for the grade inflation probability too.

Overall, the results suggest strongly that the onset of individual teacher incentives led to a decrease in student achievement and an increase of grade inflation. As mentioned before, this is consistent with incentives-related disruption in collaborative work in schools, once teachers are facing tournaments with their colleagues. This is also consistent with teaching to the test in terms of internal results, those which carry a greater weight in terms of the final result of the student in the module (70%). However, given the non-random nature of the intervention, we conduct additional robustness in the next section, before reaching our conclusions.

5 Robustness

One important test of the strength of a causal interpretation of difference-in-differences estimates concerns common trends. Indeed, if there are no interactions between treatment and other variables, as assumed for identification purposes, one would expect parallel movement between the treatment and control groups. We conduct this analysis here by considering more flexible versions of equations 1 and 2. Specifically, we allow the difference of the treatment group with respect to the control group to vary during the period prior to the intervention. If our earlier estimates are indeed capturing a causal effect, then we expect that there will be no statistically significant differences between the two groups (control and treatment) until the occurrence of the treatment.

Moreover, we also allow the effect of the education reform to vary over the *After* period. This serves as another robustness test, as it allows us to investigate any cumulative effects of the reform. In this context, the equation we estimate is as follows:

$$y_{ijt} = \alpha_0 + \alpha_1 \text{Continent}_j + \sum_{k=2003}^{2008} \delta_k I(\text{year}_t = k) + \sum_{k=2003}^{2008} \gamma_k \text{Continent}_j \times I(\text{year}_t = k) + u_{ijt}. \quad (3)$$

All variables have the same meaning as before; and $I()$ is the indicator function. The coefficients of interest are now the γ_k ($k=2003, \dots, 2008$), which will indicate any differences in the yearly effects of the treatment group with respect to the comparison year (2002). Again, we also consider specifications with region, school or school-exam fixed effects.

Table 10 presents the results based on the internal grades as the dependent variable. We find that, across all specifications, there are no differences in trends between public schools in Azores and Madeira and those in the continent. We also find that there are significant treatment effects in 2007, but not in 2008. In terms of external results - Table 11 -, we find again no evidence of different trends between the two types of schools. On the other hand, the significant negative results of public schools in the continent are concentrated in 2008 (although the point estimates increase considerably from 2006 to 2007). These results are also displayed in Figure 1.

The decomposition of the results on grade inflation are again consistent with the earlier findings. Table 12 indicates no systematic differences between the two types of schools until 2008, when grade inflation effects jump in magnitude and become statistically significant in

all cases except one. Before that, in 2007, point estimates are already typically much higher than before.

Overall, we regard these results are very supportive of our causal interpretation. Furthermore, the cumulative nature of the effects is also consistent with the cumulative nature of the reform, in the sense that the cohorts that take their exams later (in 2008 rather than in 2007) are also cohorts that have been exposed to the treatment for a longer period.

We also test the common trends assumption (and the cumulateness of the effects) in terms of the public vs private schools comparison:

$$y_{ijt} = \alpha'_0 + \alpha'_1 Public_j + \sum_{k=2003}^{2008} \delta_k I(year_t = k) + \sum_{k=2003}^{2008} \gamma'_k Public_j \times I(year_t = k) + u_{ijt}. \quad (4)$$

Again we find evidence of no statistically significant differences during the *before* period for all three variables considered (internal and external grades and grade inflation) and across the four specifications estimated for each variable - see Tables 13 and 14 (and Figure 2) for the results on internal and external marks and Table 15 for the results on grade inflation.⁹ Without exception, all point estimates in 2008 are bigger (in absolute terms) than their counterparts for 2007, although their differences are typically not statistically significant. However, we regard this additional finding about the increase of the point estimates as further evidence consistent with the cumulative effects of the reform.

We have also conducted a number of additional robustness tests. First, we considered different subsets of our original data. Specifically, we consider possible differences from the benchmark results alternatively in urban areas, in large schools or in core subjects. Second, we extended the range of data examined from first-sit results (which account for about 85% of the total number of exams) to first- and second-sits. Third, we extended the range of data considered even further, considering also resit students and those not applying for university entry. Fourth, we replicated our results using data collapsed into before and after periods, as suggested in Bertrand et al. (2004). In all cases, the qualitative results across the different specifications were unchanged and only relatively minor differences were found in terms of the quantitative findings. Finally, our early results based on the case of primary schools (Martins

⁹The only exception to this pattern is some evidence of lower external results and higher inflation in 2004 and 2005, but only in two specifications. Those point estimates are generally quite smaller than the 2007 or 2008 point estimates.

2009) also finds similar qualitative results as for secondary schools.¹⁰

6 Conclusions

There is great interest in understanding the potential of teacher incentives to improve student achievement. This paper sheds light into this question by examining the recent introduction of performance-related pay in Portugal's public schools. Our approach is based on a difference-in-differences analysis drawing on two complementary control groups. These control groups either were exposed to a lighter version of the intervention (the case of public schools in the Azores and Madeira) or were not exposed at all (the case of private schools). All students in all schools were administered the same national exams.

Our results consistently indicate that the increased focus on individual teacher performance caused a significant and sizable relative decline in student achievement, as measured by national exams. However, the decline in achievement is less sizable when considering school-level results, suggesting an increasing importance of grade inflation. This result is also supported by our triple-difference evidence. Furthermore, we find additional support for a causal interpretation of our results from our analysis of common trends, robustness to different control variables, different data subsets and different aggregation levels. The use of official population data also ensures greater reliability in terms of the external validity of our results and measurement bias.

Overall, our student achievement results are consistent with the potential for disruption in collaborative work in schools created by tournaments for promotions and prizes. On the other hand, the increase in grade inflation can be explained by the emphasis placed by the new promotion criteria on student results. While our results are negative in terms of the value of the specific reform examined here, our findings also indicate that teachers respond to incentives in a predictable way. In this context, we believe that future research should move from the question of whether performance-related pay has any effects to the narrower question of which specific performance-related pay setups generate the best results for students.

¹⁰We also consider the effects in terms of the public vs private contrast within Azores and Madeira only. The results are again very similar to our benchmark findings. However they are less precise, as there is only one private school in this subset. In addition, we also contrasted the percentage of students that applied for university entry between the treatment and the control groups. Here we found some evidence of a greater percentage of public-school students applying for university. However, it is unclear if this results from composition differences across the two groups over time or if this is an effect itself of the grade inflation documented before.

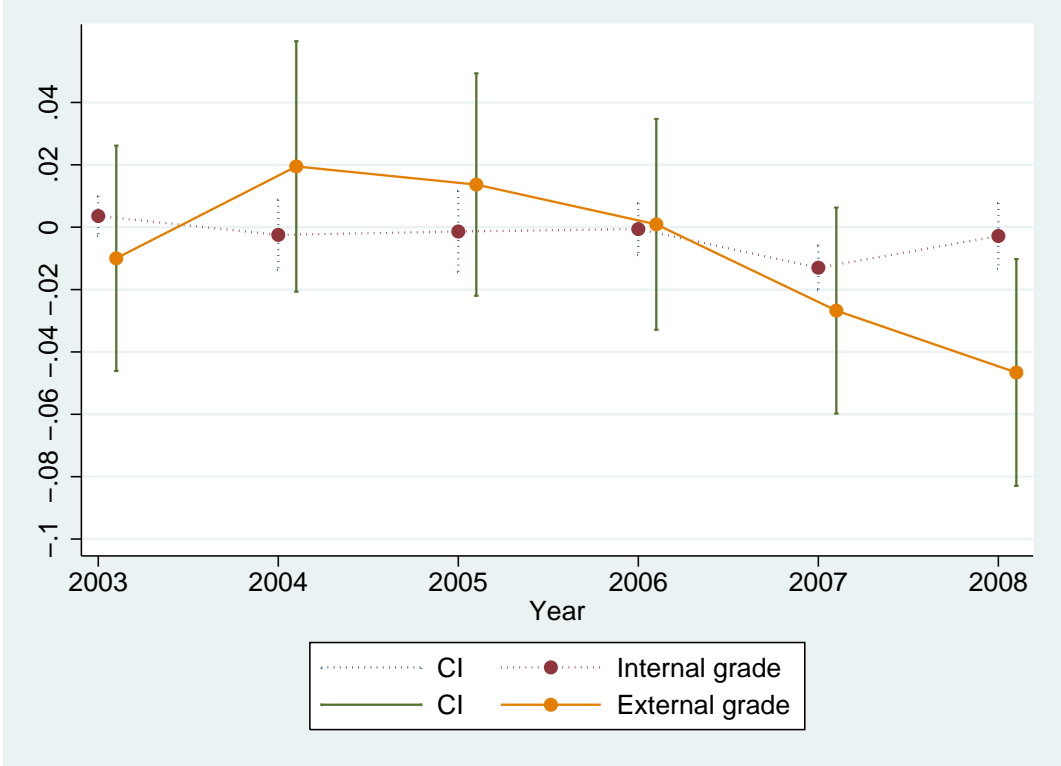
References

- Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H. & Wilson, D. (forthcoming), ‘Evaluating the impact of performance-related pay for teachers in England’, *Labour Economics* .
- Bandiera, O., Barankay, I. & Rasul, I. (2005), ‘Social preferences and the response to incentives: Evidence from personnel data’, *Quarterly Journal of Economics* **120**(3), 917–962.
- Bertrand, M., Duflo, E. & Mullainathan, S. (2004), ‘How much should we trust differences-in-differences estimates?’, *Quarterly Journal of Economics* **119**(1), 249–275.
- Eberts, R., Hollenbeck, K. & Stone, J. (2002), ‘Teacher performance incentives and student outcomes’, *Journal of Human Resources* **37**(4), 913–927.
- European Commission (2007), The education system in Portugal, Eurybase report.
- Fehr, E. & Schmidt, K. M. (1999), ‘A theory of fairness, competition, and cooperation’, *Quarterly Journal of Economics* **114**(3), 817–868.
- Figlio, D. N. & Kenny, L. W. (2007), ‘Individual teacher incentives and student performance’, *Journal of Public Economics* **91**(5-6), 901–914.
- Glewwe, P., Ilias, N. & Kremer, M. (2003), Teacher incentives, NBER Working Paper 9671.
- Jacob, B. A. & Levitt, S. D. (2003), ‘Rotten apples: An investigation of the prevalence and predictors of teacher cheating’, *Quarterly Journal of Economics* **118**(3), 843–877.
- Kane, T. J. & Staiger, D. O. (2002), ‘The promise and pitfalls of using imprecise school accountability measures’, *Journal of Economic Perspectives* **16**(4), 91–114.
- Lavy, V. (2002), ‘Evaluating the effect of teachers’ group performance incentives on pupil achievement’, *Journal of Political Economy* **110**(6), 1286–1317.
- Lavy, V. (forthcoming), ‘Performance pay and teachers effort, productivity and grading ethics’, *American Economic Review* .
- Lazear, E. P. (2000), ‘Performance pay and productivity’, *American Economic Review* **90**(5), 1346–1361.

- Lazear, E. P. (2003), 'Teacher incentives', *Swedish Economic Policy Review* **10**(2), 179–214.
- Lazear, E. P. & Rosen, S. (1981), 'Rank-order tournaments as optimum labor contracts', *Journal of Political Economy* **89**(5), 841–64.
- Martins, P. S. (2008), 'Dispersion in wage premiums and firm performance', *Economics Letters* **101**(1), 63–65.
- Martins, P. S. (2009), Teacher performance-related pay and student achievement: Evidence from primary schools, Queen Mary, University of London, mimeo.
- Martins, P. S. & Walker, I. (2006), Student achievement and university classes: Effects of attendance, size, peers, and teachers, IZA Discussion Paper 2490.
- Meyer, B. D. (1995), 'Natural and quasi-experiments in economics', *Journal of Business & Economic Statistics* **13**(2), 151–61.
- Neal, D. (1997), 'The effects of catholic secondary schooling on educational achievement', *Journal of Labor Economics* **15**(1), 98–123.
- OECD (2001), 'Education at a glance 2001', OECD Directorate for Education, Paris.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (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.

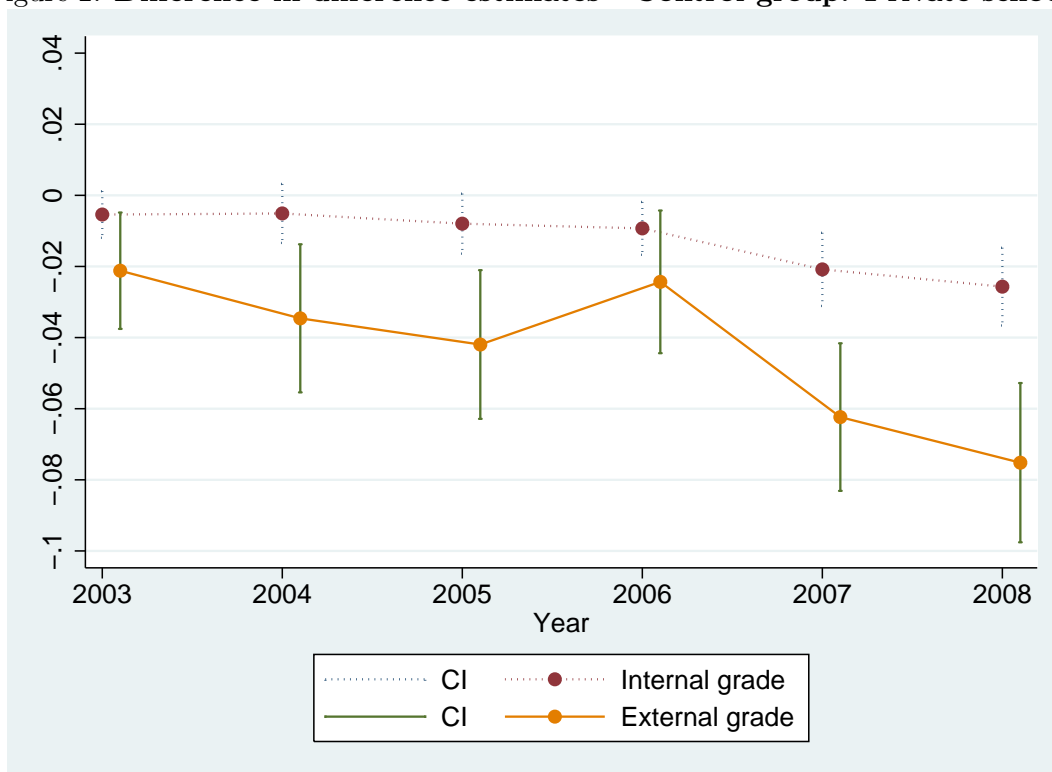
Figures

Figure 1: Difference-in-difference estimates - Control group: Azores and Madeira



Notes: Author's estimates based on JNE data and reported in more detail in Column C of Tables 10 and 11.

Figure 2: Difference-in-difference estimates - Control group: Private schools



Notes: Author's estimates based on JNE data and reported in more detail in Column C of Tables 13 and 14.

Tables

Table 1: **Descriptive Statistics - All Exams**

Variable	Mean	Std. Dev.	Min.	Max.
Internal Exam	13.322	2.598	10	20
External Exam	10.66	3.992	0.1	20
Internal - External Exam	2.662	3.148	-10	18.9
Log Internal Exam	2.571	0.19	2.303	2.996
Log External Exam	2.276	0.47	-2.303	2.996
Public	0.888	0.315	0	1
Continent	0.949	0.221	0	1
No. Exams	651.440	390.932	1	3152
2003	0.166	0.372	0	1
2004	0.121	0.327	0	1
2005	0.135	0.342	0	1
2006	0.164	0.37	0	1
2007	0.112	0.316	0	1
2008	0.115	0.319	0	1

Notes: Author's calculations based on *Jurí Nacional de Exames* data. The internal (external) exam refers to the mark obtained by each student in each module by the school (national exam). 'Public' and 'Continent' are dummy variables which are equal to one for students in public schools or schools located in mainland Portugal, respectively. There are 1411246 observations in each variable.

Table 2: **Descriptive Statistics - All Schools**

Variable	Mean	Std. Dev.	Min.	Max.
Internal Exam	13.055	0.687	10.333	16.356
External Exam	10.123	1.308	5.467	15.72
Internal - External Exam	2.932	1.028	-1.564	6.653
Log Internal Exam	2.55	0.05	2.334	2.777
Log External Exam	2.212	0.154	1.569	2.735
Public	0.833	0.373	0	1
Continent	0.945	0.228	0	1
No. Exams	417.287	313.339	1	3152
2003	0.143	0.35	0	1
2004	0.143	0.35	0	1
2005	0.143	0.35	0	1
2006	0.143	0.35	0	1
2007	0.143	0.35	0	1
2008	0.143	0.35	0	1

Notes: Author's calculations based on *Jurí Nacional de Exames* data. The internal (external) exam refers to the mark obtained by each student in each module by the school (national exam). 'Public' and 'Continent' are dummy variables which are equal to one for students in public schools or schools located in mainland Portugal, respectively. There are 3941 observations in each variable.

Table 3: **Effects on internal grades - Islands control group**

	(1)	(2)	(3)	(4)
After	.012 (.006)*	.015 (.006)**	.010 (.005)*	.021 (.006)***
Continent	.003 (.006)			
Continent-After	-.009 (.007)	-.008 (.007)	-.008 (.005)	-.010 (.006)
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1047472	1047472	1047472	664637
R^2	.0001	.004	.022	.052

Notes: Dependent variable is the log of the school-level grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 4: Effects on external grades - Islands control group

	(1)	(2)	(3)	(4)
After	.033 (.012)***	.047 (.014)***	.028 (.013)**	.085 (.016)***
Continent	.114 (.019)***			
Continent-After	-.040 (.013)***	-.042 (.015)***	-.042 (.013)***	-.054 (.016)***
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1046910	1046910	1046910	664215
R^2	.003	.016	.045	.123

Notes: Dependent variable is the log of the national exam grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 5: Effects on grade inflation - Islands control group

	(1)	(2)	(3)	(4)
After	-.036 (.111)	-.119 (.121)	-.026 (.118)	-.249 (.143)*
Continent	-.914 (.129)***			
Continent-After	.217 (.114)*	.233 (.123)*	.244 (.119)**	.246 (.145)*
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1047472	1047472	1047472	664637
R^2	.004	.02	.053	.162

Notes: Dependent variable is the difference between the internal (school) grade and the external (national exam) grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 6: Effects on grade inflation probability - Islands control group

	(1)	(2)	(3)	(4)
After	.005 (.009)	-.002 (.009)	.001 (.009)	-.009 (.012)
Continent	-.068 (.009)***			
Continent-After	.018 (.010)*	.019 (.010)*	.022 (.010)**	.016 (.012)
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1047472	1047472	1047472	664637
R^2	.002	.008	.025	.079

Notes: Dependent variable is one if the internal (school) grade is greater than the external (national exam) grade, for each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 7: Effects on internal grades - Private schools control group

	(1)	(2)	(3)	(4)
After	.026 (.005)***	.032 (.006)***	.019 (.005)***	.027 (.005)***
Public	-.034 (.010)***	-.034 (.010)***		
Public-After	-.022 (.005)***	-.025 (.006)***	-.019 (.005)***	-.017 (.006)***
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1120030	1120030	1120030	713828
R^2	.005	.01	.038	.071

Notes: Dependent variable is the log of the school-level grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 8: **Effects on external grades - Private schools control group**

	(1)	(2)	(3)	(4)
After	.052 (.012)***	.066 (.014)***	.031 (.010)***	.074 (.011)***
Public	-.028 (.017)*	-.012 (.019)		
Public-After	-.058 (.012)***	-.065 (.015)***	-.048 (.011)***	-.046 (.012)***
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1119430	1119430	1119430	713376
R^2	.001	.013	.05	.125

Notes: Dependent variable is the log of the national exam grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 9: **Effects on grade inflation - Private schools control group**

	(1)	(2)	(3)	(4)
After	-.067 (.098)	-.125 (.099)	.035 (.086)	-.214 (.088)**
Public	-.166 (.130)	-.285 (.147)*		
Public-After	.248 (.102)**	.270 (.108)**	.193 (.091)**	.227 (.093)**
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1120030	1120030	1120030	713828
R^2	.0007	.014	.054	.159

Notes: Dependent variable is the difference between the internal (school) grade and the external (national exam) grade of each student in each exam in each year. Dummy *After* is one for 2007 and 2008 only. Data used: 2002 to 2008, except for 2006. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 10: **Effects on internal grades - Islands control group**

	(1)	(2)	(3)	(4)
2003	-.008 (.006)	-.002 (.005)	-.004 (.004)	-.0001 (.004)
2004	.0009 (.008)	.007 (.010)	.001 (.007)	.013 (.007)**
2005	-.002 (.008)	.003 (.010)	-.001 (.008)	.005 (.007)
2006	.013 (.005)**	.011 (.006)**	.012 (.005)**	.019 (.006)***
2007	.015 (.005)***	.021 (.005)***	.013 (.004)***	.031 (.006)***
2008	.004 (.007)	.012 (.008)	.002 (.007)	.023 (.009)***
Continent	.0004 (.006)			
Continent-2003	.009 (.006)	.005 (.005)	.004 (.004)	.002 (.004)
Continent-2004	-.0004 (.008)	-.0004 (.009)	-.002 (.007)	-.007 (.007)
Continent-2005	.001 (.009)	.0009 (.010)	-.001 (.008)	-.002 (.007)
Continent-2006	.0005 (.005)	.003 (.006)	-.0006 (.005)	-.002 (.006)
Continent-2007	-.011 (.005)**	-.011 (.005)**	-.013 (.004)***	-.015 (.006)**
Continent-2008	-.0009 (.007)	-.003 (.008)	-.003 (.007)	-.009 (.009)
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1253690	1253690	1253690	800521
R^2	.0007	.004	.022	.053

Notes: Dependent variable is the log of the school-level grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 11: Effects on external grades - Islands control group

	(1)	(2)	(3)	(4)
2003	.005 (.021)	.019 (.018)	.003 (.022)	-.040 (.023)*
2004	-.058 (.020)***	-.033 (.019)*	-.076 (.025)***	-.071 (.029)**
2005	-.007 (.018)	.011 (.017)	-.018 (.022)	-.033 (.028)
2006	-.059 (.022)***	-.063 (.030)**	-.064 (.021)***	-.066 (.028)**
2007	-.040 (.018)**	-.018 (.018)	-.067 (.020)***	-.024 (.024)
2008	.088 (.020)***	.117 (.022)***	.059 (.023)**	.116 (.028)***
Continent	.108 (.027)***			
Continent-2003	-.007 (.021)	-.015 (.018)	-.010 (.023)	.010 (.024)
Continent-2004	.019 (.020)	.015 (.019)	.019 (.025)	.025 (.030)
Continent-2005	.019 (.018)	.016 (.017)	.014 (.022)	.024 (.028)
Continent-2006	.007 (.023)	.014 (.031)	.0009 (.021)	-.0007 (.028)
Continent-2007	-.025 (.019)	-.026 (.019)	-.027 (.021)	-.031 (.024)
Continent-2008	-.048 (.021)**	-.059 (.022)***	-.047 (.023)**	-.055 (.028)*
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1253044	1253044	1253044	800037
R^2	.008	.02	.049	.141

Notes: Dependent variable is the log of the national exam grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 12: **Effects on grade inflation - Islands control group**

	(1)	(2)	(3)	(4)
2003	-.078 (.182)	-.145 (.190)	-.011 (.206)	.416 (.188)**
2004	.483 (.160)***	.348 (.129)***	.654 (.205)***	.761 (.195)***
2005	-.033 (.158)	-.135 (.136)	.072 (.186)	.245 (.173)
2006	.737 (.185)***	.755 (.226)***	.774 (.164)***	.937 (.229)***
2007	.610 (.159)***	.488 (.162)***	.810 (.174)***	.781 (.218)***
2008	-.581 (.170)***	-.746 (.181)***	-.360 (.193)*	-.492 (.212)**
Continent	-.916 (.182)***			
Continent-2003	.168 (.184)	.208 (.192)	.133 (.206)	-.026 (.190)
Continent-2004	-.078 (.163)	-.056 (.129)	-.119 (.199)	-.216 (.193)
Continent-2005	-.124 (.162)	-.105 (.137)	-.115 (.183)	-.207 (.173)
Continent-2006	-.003 (.188)	-.040 (.229)	.033 (.166)	-.004 (.231)
Continent-2007	.119 (.164)	.130 (.162)	.112 (.171)	.048 (.215)
Continent-2008	.368 (.174)**	.432 (.183)**	.335 (.188)*	.277 (.208)
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1253690	1253690	1253690	800521
R^2	.017	.031	.062	.19

Notes: Dependent variable is the difference between the internal (school) grade and the external (national exam) grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 13: Effects on internal grades - Private schools control group - Year decomposition

	(1)	(2)	(3)	(4)
2003	.006 (.004)	.006 (.004)	.005 (.004)	.006 (.004)
2004	.011 (.006)*	.011 (.006)*	.002 (.005)	.004 (.006)
2005	.010 (.006)*	.009 (.006)	.004 (.005)	.006 (.006)
2006	.028 (.005)***	.028 (.006)***	.020 (.004)***	.024 (.005)***
2007	.036 (.007)***	.036 (.007)***	.019 (.005)***	.031 (.006)***
2008	.040 (.008)***	.040 (.008)***	.023 (.006)***	.032 (.007)***
Public	-.034 (.010)***	-.031 (.009)***		
Public-2003	-.004 (.004)	-.003 (.004)	-.005 (.004)	-.005 (.005)
Public-2004	-.004 (.006)	-.004 (.006)	-.005 (.005)	-.0005 (.006)
Public-2005	-.005 (.005)	-.005 (.005)	-.008 (.005)	-.004 (.006)
Public-2006	-.013 (.006)**	-.014 (.006)**	-.009 (.005)**	-.007 (.006)
Public-2007	-.025 (.007)***	-.026 (.007)***	-.021 (.006)***	-.018 (.007)***
Public-2008	-.030 (.008)***	-.031 (.008)***	-.026 (.007)***	-.020 (.008)**
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1339542	1339542	1339542	859143
R^2	.008	.01	.037	.072

Notes: Dependent variable is the log of the school-level grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 14: **Effects on external grades - Private schools control group - Year decomposition**

	(1)	(2)	(3)	(4)
2003	.009 (.012)	.009 (.012)	.014 (.010)	-.009 (.012)
2004	.001 (.016)	-.002 (.016)	-.025 (.012)**	-.010 (.018)
2005	.053 (.014)***	.051 (.014)***	.036 (.012)***	.039 (.015)***
2006	-.017 (.016)	-.019 (.016)	-.040 (.012)***	-.041 (.014)***
2007	.026 (.019)	.019 (.018)	-.034 (.012)***	.004 (.014)
2008	.141 (.020)***	.135 (.019)***	.084 (.013)***	.134 (.014)***
Public	-.024 (.021)	-.0007 (.020)		
Public-2003	-.005 (.013)	-.006 (.012)	-.021 (.010)**	-.023 (.013)*
Public-2004	-.020 (.016)	-.020 (.016)	-.035 (.013)***	-.040 (.018)**
Public-2005	-.026 (.015)*	-.026 (.014)*	-.042 (.013)***	-.051 (.016)***
Public-2006	-.031 (.017)*	-.031 (.016)*	-.024 (.013)*	-.026 (.015)*
Public-2007	-.070 (.020)***	-.070 (.019)***	-.062 (.013)***	-.062 (.015)***
Public-2008	-.081 (.021)***	-.082 (.020)***	-.075 (.014)***	-.075 (.015)***
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1338860	1338860	1338860	858631
R^2	.01	.018	.053	.142

Notes: Dependent variable is the log of the national exam grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.

Table 15: Effects on grade inflation - Private schools control group - Year decomposition

	(1)	(2)	(3)	(4)
2003	.028 (.104)	.023 (.100)	-.028 (.083)	.243 (.099)**
2004	.203 (.102)**	.222 (.099)**	.317 (.084)***	.237 (.130)*
2005	-.372 (.110)***	-.362 (.109)***	-.294 (.100)***	-.335 (.124)***
2006	.639 (.115)***	.657 (.112)***	.742 (.110)***	.823 (.138)***
2007	.356 (.131)***	.411 (.126)***	.679 (.109)***	.541 (.129)***
2008	-.713 (.134)***	-.668 (.129)***	-.420 (.111)***	-.712 (.121)***
Public	-.196 (.152)	-.354 (.135)***		
Public-2003	.037 (.105)	.051 (.102)	.151 (.088)*	.144 (.105)
Public-2004	.106 (.102)	.111 (.099)	.224 (.088)**	.302 (.136)**
Public-2005	.143 (.114)	.152 (.112)	.256 (.106)**	.368 (.129)***
Public-2006	.076 (.120)	.066 (.117)	.066 (.116)	.108 (.145)
Public-2007	.269 (.137)**	.254 (.130)*	.249 (.113)**	.280 (.133)**
Public-2008	.402 (.143)***	.398 (.137)***	.401 (.115)***	.489 (.125)***
Region fixed effects		X		
School fixed effects			X	
School-exam fixed effects				X
Obs.	1339542	1339542	1339542	859143
R^2	.015	.026	.063	.187

Notes: Dependent variable is the difference between the internal (school) grade and the external (national exam) grade of each student in each exam in each year. Columns 2-4 includes a control for the number of exams taken in each school in each year. Robust standard errors, allowing for clustering at the school level. Significance levels: *: 0.05; **: 0.01; ***: 0.001.