Teacher Performance Pay and Student Learning: Evidence from a Nationwide Program in Peru*

Cristina Bellés Obrero†   María Lombardi‡

JOB MARKET PAPER

LINK TO THE LATEST VERSION

Abstract

We study a nationwide teacher pay-for-performance program implemented in public secondary schools in Peru in 2015, and examine its impact on student performance. The program takes the form of a tournament, awarding a bonus of over a month's salary to the principal and every teacher from schools in the top 20 percent within a group of comparable schools. Exploiting the fact that the main performance measure used to rank schools in this tournament is the average score of 8^{th} graders in a 2015 standardized test, we perform a difference-in-differences estimation comparing changes in the internal grades of 8^{th} graders before and after the incentive was introduced to those of 9^{th} graders from the same school. We find that the teacher pay-for-performance program had a precisely estimated zero effect on student achievement, allowing us to reject impacts greater than 0.017 standard deviations, well below those previously found in the literature. We provide evidence against a series of potential explanations, and argue that this zero effect can be explained by some of the program's characteristics, which may have hindered teachers' ability to improve the incentivized outcome or infer their probability of winning.

JEL Classifications: I21, M52, J4
Keywords: education, teachers, incentives, compensation, Peru.

*We are especially grateful to Libertad González, Gianmarco León, Stephan Litschig and Alessandro Tarozzi. We also thank Ruben Enikopolov, Rosa Ferrer, Ana Figueiredo, Niklas Heusch, Ricardo Maertens, Ana Tur-Prats, Tomás Williams, and participants at EMCON, NEUDC and UPF LPD Seminar for their helpful comments. We are also grateful to the Peruvian Ministry of Education for providing and assisting with the data used in the study. The views expressed herein are those of the authors and do not necessarily reflect those of the Peruvian Government.

†cristina.belles@upf.edu. PhD candidate, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain. Bellés Obrero aknowledges support from the Spanish Ministry of Economy and Competitiveness.

‡maria.lombardi@upf.edu (corresponding author). PhD candidate, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain. Lombardi acknowledges support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Predoctoral Programme.

1 Introduction

Teacher quality is one of the key factors determining student achievement (Hanushek and Rivkin, 2010; Rockoff, 2004). Individuals exposed to better teachers not only perform better in school, but are also more likely to attend college and earn higher salaries (Chetty et al., 2014). However, the payment schemes in most educational systems do not provide adequate incentives for excellence in teaching. With relatively flat salary progression, promotion policies rigidly linked to seniority, and lifetime job tenure, these types of compensation policies might discourage high skilled individuals from taking on the teaching profession, and create weak incentives for existing teachers to exert high levels of effort (Bruns et al., 2011). In an attempt to increase teacher motivation, accountability, effort, and ultimately student learning, academics and policymakers have proposed tying teachers' pay to their students' performance. Pay-for-performance programs in education have been implemented in high income countries like United States, England, the Netherlands and Israel,[1] as well as in developing countries such as India, Kenya, China, Chile, Brazil, Mexico, and more recently in Peru. However, the evidence on the effectiveness of teacher incentives is scant and inconclusive; only part of these programs have been rigorously evaluated, and those that have been studied often differ in their conclusions.

This paper studies the impact of *Bono Escuela* (BE) on student achievement. BE is a nationwide teacher pay-for-performance program implemented in 2015 in public secondary schools in Peru. The program takes the form of a rank-order tournament in which all Peruvian public secondary schools compete within a group of comparable schools on the basis of their annual performance. Every teacher and the principal in schools ranked in the top 20% within their BE group obtain a fixed payment amounting to over a month's salary. The incentives provided by the BE are collective (at the school-level), as all teachers are rewarded if their school wins, although the main performance measure used to rank schools is their average score in the 2015 nationwide math and language standardized tests, taken only by 8^{th} graders. This feature of the program, which we exploit in our identification, implies that a school's probability of obtaining the bonus hinges on the achievement of 8^{th} grade students in 2015. Our estimation relies on a novel administrative database collected by the Peruvian Ministry of Education, which covers the universe of students in 2013-2015 and contains annual information on the grades that students receive from their teachers in each subject (their "internal grades").[2] Importantly,

[1] An exhaustive list of OECD countries with teacher pay-for-performance programs is provided in OECD Education at a Glance 2011, available at https://www.oecd.org/education/skills-beyond-school/48631582.pdf

[2] We borrow this terminology from Calsamiglia and Loviglio (2016).

teachers' grading tactics should not be influenced by the incentive, since internal grades have no direct impact on a school's BE score. We provide ample evidence that internal grades are correlated with standardized measures of learning, and show that the same teacher typically grades students from parallel classes differently, suggesting that grading on a curve is not the norm in Peruvian secondary schools. The availability of achievement measures for students in all grades allows us to compare changes in the internal grades of 8^{th} graders to those of 9^{th} grade students attending the same school, before and after the incentive was introduced, providing difference-in-difference estimates of the effect of BE on student achievement.

While providing teachers with extrinsic monetary incentives could encourage them to exert more effort, positively impacting student performance, these incentives might yield no improvement if the incentives are not large enough, not understood, or if teachers do not know how to increase student achievement, for example. Teacher incentives might also be ineffective, or even detrimental to student learning, if they lead teachers to engage in undesirable practices such as targeting topics likely to be tested, coaching students on test-taking strategies, or cheating.[3] Student learning could also decrease if the program crowds out teachers' intrinsic motivation.[4] Since a school's probability of obtaining the BE bonus depends on the effort of many of its teachers, the program has the potential of inducing teacher free-riding (Holmstrom, 1982), thus lowering its impact on student learning. However, rewarding the entire school might have the benefit of promoting higher cooperation and monitoring among teachers (Kandel and Lazear, 1992; Kandori, 1992).

We find that the program had no impact on students' math and language internal grades. Our coefficients are precisely estimated, allowing us to reject effects larger than 0.017 standard deviations (SD) in both math and language, well below the treatment effects found in the existing literature (around 0.15-0.25 SD). Furthermore, when separately examining the impact of the program in each of the 395 groups in which schools compete, we find zero average effects in the majority of cases, providing additional evidence of the null effect of BE on student achievement. Using data on the

[3]This type of behavior is consistent with models of multi-tasking (Holmstrom and Milgrom, 1991; Baker, 1992, 2002), and has been reported in several studies on teacher incentives such as Jacob and Levitt, 2003, Figlio and Winicki, 2005, Figlio, 2006, Glewwe et al., 2010, and Behrman et al., 2015, to name a few. Although these actions could still improve the performance of students whose teachers were devoting little time to effective teaching, they might not affect student learning, or even harm it if they crowd-out effective instruction time (Koretz, 2002; Neal, 2011).

[4]Previous research shows that monetary incentives produce not only a price effect, making the incentivized behavior more attractive, but also a psychological effect that can crowd out the former. There is evidence of this behavior, albeit in a different context, in the studies of Gneezy and Rustichini (2000a), Gneezy and Rustichini (2000b) and Gneezy et al. (2011).

overlap of teachers in 8^{th} and 9^{th} grade, we assess whether the BE generated improvements in student achievement in our comparison group, and discard the existence of any spillovers which could bias our estimates. Our null average treatment effects could be masking the fact that teachers from some schools might be more incentivized than other, due to the tournament nature of the BE. In particular, the incentive could be impacting schools closer to the margin, anpd leaving sure-winners and sure-losers unaffected (Contreras and Rau, 2012). We explore whether there are differential effects across these and other dimensions of the incentive, and do not find evidence of heterogeneous effects.

Why was student learning unaffected by the BE program? In order to answer this question, we carried out an online survey on a sample of public secondary school teachers regarding the 2015 BE. We provide suggestive evidence that the null effect is not a result of the size or collective nature of the incentive, or driven by teachers being uninformed about the BE, or only focusing on increasing standardized test scores -the incentivized outcome- without influencing their students' learning in a meaningful way. We put forth a few reasons why the program may have had a null effect. Firstly, certain features of the standardized test linked to the bonus might have hampered teachers' ability to boost student performance in terms of this measure, potentially discouraging teachers from exerting higher effort. Given that students were tested for the first time in 2015, teachers might not have known what pedagogical practices result in higher test scores. The fact that students have no stakes in these evaluations might also have played a role in weakening the mapping between teachers' effort and their chances of winning the bonus. Secondly, the incentive might have been diluted if schools were uncertain about their potential ranking within the group of schools they were competing against. Given that they had no prior experience with the standardized test tied to BE, this is not unlikely. Finally, we argue that teachers might not have had enough time to react to the incentive. The future analysis of the 2016 wave of the BE, for which some of these issues will be alleviated, will also allow us to better pin-down these channels.

Our paper relates to the literature on teacher pay-for-performance, particularly in the context of other developing countries. Although a few studies find positive and significant effects on student learning, the literature reveals mixed results. Using a randomized controlled trial in rural schools in the Indian state of Andhra Pradesh, Muralidharan and Sundararaman (2011) study the effect of providing individual and collective monetary incentives to teachers based on students' test score improvements. The incentive had a significant and sizable effect on students' standardized test scores and a positive impact on other subjects not targeted by the incentive. The experimental study of Glewwe et al. (2010) evaluates a collective teacher incentive program in Kenya, and finds that although the program yielded a positive effect on students' test scores in the exams tied to the incentive, it had no impact on non-incentivized exams covering similar topics.[5] Behrman et al. (2015) implement a randomized controlled trial in a sample of Mexican high schools, providing monetary incentives to teachers and/or students based on the latter's performance in math tests with very low stakes. The authors find that, while providing monetary incentives to teachers had no impact on students' math test scores, there was a significant increase in student performance when students themselves were incentivized. The effects were larger when both students and teachers were given incentives.[6] The closest paper to ours is that of Contreras and Rau (2012), who examine the impact of a scaled-up program in Chile. Using matching and difference-in-difference techniques, they find that public school students performed significantly better in math and language as compared to students in private schools, which were not eligible for the bonus.

In the context of a high income country, the quasi-experimental studies of Lavy (2002) and Lavy (2009) in Israeli high schools examine a collective and an individual teacher incentive program, respectively, and find positive and significant impacts in different measures of student performance tied to the incentives. In a follow-up paper, Lavy (2015) finds that ten years after the pay-for-performance program examined in Lavy (2009), students from treated schools exhibited significantly higher level of schooling attainment and higher wages. Fryer (2013), and Goodman and Turner (2013) independently analyze a randomized controlled experiment in over 200 New York City public schools where schools meeting their performance target could earn a lump sum payment, which they could distribute at their own discretion. Both studies find no evidence of increased student attainment or changes in students' or teachers' behavior. Finally, Springer et al. (2010) conducted a three-year study in the Metropolitan Nashville School System in which math teachers were economically incentivized for large gains on standardized tests, and find a positive effect only among teachers instructing the same set of

[5]Although direct observation of teachers in Muralidharan and Sundararaman (2011) shows no impact of the teacher incentive on classroom processes, or student and teacher attendance, teachers in treatment schools were more likely to report having assigned extra homework, classwork and practice tests, conducting extra classes, and paying special attention to weaker students. External observers in Glewwe et al. (2010) also found no changes in teacher attendance, homework assignment or pedagogy. However, the principals of treated schools were more likely to report that their teachers offered extra prep classes, suggesting that teachers' efforts might have narrowly targeted the incentivized outcome.

[6]Consistent with the authors' findings, incentivized students reported exerting higher effort in preparing for the exam. Self-reported behavior of teachers is not as compatible with the results, since teachers in all three treatment arms were more likely to report having prepared their students for the test, both inside and outside of class.

students in multiple subjects.

Performance pay schemes have traditionally been examined in the context of organizations.[7] There have been several studies examining the causal effect of linking managerial pay to overall firm performance (Groves et al., 1994, Chevalier and Ellison, 1997, and Oyer, 1998, among others) or to the productivity of bottom-tier workers (Bandiera et al., 2007). Other papers have focused instead on the impact of different payment schemes on worker productivity (e.g., Bandiera et al., 2005 and Bandiera et al., 2013). The results of our study can be informative for this other stream of the literature as well.

One contribution of our paper is that we examine whether teacher pay-for-performance can work in the context of a scaled-up, national intervention. Except for Contreras and Rau (2012), all the other studies in this literature tackle this question using a randomized controlled trial, in which the scale is necessarily smaller. Muralidharan and Sundararaman (2011) implement a pay-for-performance program in 200 schools, Glewwe et al. (2010) in 50 schools, and Behrman et al. (2015) in 40 schools. In contrast, the BE program reached more than 8,000 schools across Peru, providing incentives to roughly 81,000 teachers, responsible for instructing 70% of Peruvian students in the 8^{th} grade. While these experimental studies make important contributions towards understanding whether teacher pay-for-performance can increase student achievement, they face external validity issues, as in any randomized controlled trial (Deaton and Cartwright, 2016), making their findings not necessarily generalizable to a large-scale program. This notion is put forward in Banerjee et al. (2016), where a successful educational intervention led by a NGO did not yield the same initial impact when it was scaled-up and implemented within the existing educational system. Budgetary constraints (Kerwin and Thornton, 2015) and opposition from teacher unions (Bruns and Luque, 2015; Mizala and Schneider, 2014) make several aspects of these types of interventions unfeasible in a nationwide program. For example, students in Muralidharan and Sundararaman (2011) were evaluated at baseline, and their teachers received feedback on their performance in each question. Testing students so often and providing such detailed feedback to their teachers might be too costly to implement on a national scale. It is thus crucial for policymakers to better understand the role played by the features of teacher pay-for-performance programs. While we cannot fully tease out which characteristics of the BE contributed to its null impact, we provide some suggestive evidence, hopefully shedding more light

[7]See Prendergast (1999) for a general review of the early empirical evidence on the provision of incentives in firms.

on this discussion.

Another novelty of our study is its use of a measure of student achievement that captures the skills of students which are targeted by the program without being directly incentivized. Since the BE bonus is linked to standardized test scores and not to internal grades, teachers' stakes in our outcome variable are not modified by the incentive.[8] An identification strategy relying on standardized test scores (or other incentivized indicators) as an outcome cannot fully disentangle whether improvements in students' performance are the consequence of higher learning or the results of short-term strategies fostering high test scores (Neal, 2011). The importance of this issue is highlighted by the results from Glewwe et al. (2010), who find that while students performed better in the tests used to award the bonus, there was no effect in their performance in an alternative exam not linked to the incentive. With the exception of the latter, all the other papers in this literature assess student achievement using measures of learning which are directly targeted by the incentive.[9] While it would also be interesting to analyze the impact of BE on standardized test scores, it is not possible because 2015 was the first year in which these tests were applied in secondary schools, and there is no appropriate comparison group. Using internal grades as our outcome has the advantage of capturing students' performance without directly influencing teachers' probability of obtaining the bonus. While this measure might have some shortcomings, for instance if teachers assign grades on a relative basis, we report the results from multiple tests alleviating this concern.

The paper is organized as follows. Section 2 describes the educational system in Peru and the Peruvian teacher pay-for-performance program. Section 3 discusses our strategy for estimating the effect of teacher incentives on student performance, and Section 4 describes the data. Section 5 presents our main results, and Section 6 provides evidence on the validity of our identification assumption. Section 7 discusses the potential reasons for our findings, and Section 8 concludes.

[8]Students' internal grades in Peru are completely independent of standardized test scores. For one, standardized tests are graded after the end of the school year, and students' individual scores are never reported.

[9]Behrman et al. (2015) and Contreras and Rau (2012) measure achievement using students' scores in the standardized evaluations tied to the bonus. The studies of Fryer (2013) and Goodman and Turner (2013) use several measures of student performance linked to the incentive (scores in state tests, graduation rates, credits earned, etc.), as do Lavy (2002) and Lavy (2009) (average score and pass rates in matriculation exams, and school dropout rates). In an attempt to overcome this issue, Muralidharan and Sundararaman (2011) designed the standardized test to include mechanical and conceptual questions; while performance in the former can be easily affected by a teacher coaching his students for the test, conceptual questions are harder to influence using these types of tactics.

2 Secondary Schooling in Peru and the BE Program

2.1 Secondary Schooling in Peru

Compulsory schooling is 12 years in Peru, and is composed of initial, primary and secondary schooling, lasting one, six, and five years. Students in public secondary schools have seven hours of instruction a day, although the Ministry of Education has been gradually implementing nine-hour school days, currently reaching 18% of all public secondary schools. While a significant portion of the student body attends private secondary schools, public institutions dominate by far. In 2014, for instance, 63% of high schools were publicly run, instructing 76% of all secondary students. Over the last decade there have been significant improvements in secondary school coverage, with enrollment rising from 71% of individuals in secondary school age (12-16) in 2005 to 83% in 2014. Despite these improvements, enrollment is still far from universal. Moreover, a very high portion of students attending high school do not possess the minimum required levels of knowledge. In the 2012 round of OECD's Programme for International Student Assessment (PISA) evaluating 15-year-old students, Peru was the lowest scoring country out of 65 in all three tested subjects. In particular, 75%, 60% and 69% of Peruvian students had low achievement in math, reading and science, respectively.

Public school teachers in Peru can be either civil servants or contract teachers. Salaries for the former are divided into eight pay scales, with a monthly salary of 1451 soles ($\approx$ 439 dollars) in the lowest scale in 2015, and a salary of 3773 soles ($\approx$ 1142 dollars) in the highest.[10] Contract teachers, on the other hand, received a fixed monthly payment of 1244 soles ($\approx$ 370 dollars).[11] There were approximately 120,000 public secondary school teachers in 2014, one third of which were contract teachers. The average secondary school teacher in public schools received a monthly salary of only 1469 soles, roughly 444 dollars.[12] The working week for public secondary school teachers in 2015 consisted of 26 hours, 24 of which were to be spent teaching.[13] However, as reported in a nationally representative teacher survey at the end of 2014, teachers spent an average of 12 hours a week performing other activities outside their official working hours, such as preparing class materials or attending parent-

[10]Throughout this study we use a conversion rate of 3.31 soles per dollar.

[11]Further details on teachers' salaries and pay scales are provided by the Ministry of Education in http://www.minedu.gob.pe/reforma-magisterial/remuneraciones-beneficios.php, last accessed August 16, 2016.

[12]We calculated the average monthly salary of public secondary school teachers in Peru using the Ministry of Education's pay scales and the type of contract and category reported by a nationally representative sample of secondary school teachers in a survey conducted by the Ministry of Education at the end of 2014 (*Encuesta Nacional de Docentes*).

[13]In 2015, the working week was expanded by two (paid) hours, which are meant to be spent performing activities outside the classroom, namely preparing materials for class, assisting students who fall behind, providing orientation to parents, etc. In 2016, an extra two hours were added, reaching a total of 30 working hours a week.

teacher conferences. Furthermore, 15% of secondary school teachers taught in more than one school, and 28% complemented their salary with another type of job.

According to this same survey, 52% of public secondary school teachers had a university degree, 45% obtained their teaching degree in a tertiary institution, and the remaining 3% had another type of degree, or no degree at all. As compared to Peruvian workers with similar qualifications, and teachers in comparable countries, Peruvian teachers are poorly paid. A study on Latin American teachers' salaries in 2010 shows that adjusting for the number of hours worked, Peruvian teachers made 10% less than other Peruvian professional workers with similar education (Bruns and Luque, 2015). In comparison to individuals with similar qualifications, teachers in Peru were paid relatively worse than in Mexico, Honduras, El Salvador, Costa Rica, Uruguay and Chile, but relatively better than than in Panama, Brazil and Nicaragua.[14] While Peruvian teachers are poorly paid, absenteeism is quite low. Around the time in which BE was implemented, teacher absenteeism was below 7% in public schools around the country.[15]

2.2 The BE Program

In 2013, the Peruvian Ministry of Education launched *Bono Escuela* (BE), a nationwide teacher pay-for-performance program in public schools. The program was first implemented in primary schools, and extended to secondary schools in 2015. Secondary schools, the focus of this paper, were only included in BE starting 2015 because Peru's census standardized tests (*Evaluación Censal de Estudiantes*, henceforth ECE), one of the key indicators used for the BE, were not implemented in secondary schools until 2015.[16] The ECE is an annual low-stakes test designed by the Peruvian Ministry of Education, in which students from different grades in practically all private and public schools are

[14] Mizala and Ñopo (2016) examine the patterns of teacher pay in several Latin American countries in an earlier period (1997-2007), and find that teachers in Nicaragua and Peru were the most underpaid relative to their nationals working as professionals or technicians.

[15] Around April 2015, the Ministry of Education launched *Semaforo Escuela*, a program in which trained enumerators make periodic visits to public schools, and register information on teacher, student, and director absenteeism, among other things. Further details are available at `http://www.minedu.gob.pe/semaforo-escuela/`. In a 2006 study, teacher absenteeism in Peruvian public schools was found to be higher, around 11% (Chaudhury et al., 2006).

[16] We do not examine the effect of the primary school BE program due to identification issues related to the timing of the program's implementation. The BE was first announced by the president of Peru in July 2014, although the corresponding regulation only came out in October of that year. The 2013 edition was implemented retroactively (i.e., after the 2013 ECE test had been taken), in an attempt to boost the program's credibility. In the case of 2014, it is unclear whether schools knew about the program before taking the ECE in November, since the BE was broadly announced four months before the test, but its regulation only came out one month before.

tested on their basic competencies in math and language at the end of the school year.[17] In secondary schools, only 8^{th} graders are tested. The ECE is implemented by the Peruvian National Statistics Institute (INEI), which trains independent enumerators for this task. Since the main goal of the ECE is to track the evolution of student learning throughout the country and help shape educational policies, school average scores are reported to school district governments, schools and parents.

Besides being an informational tool for the Ministry of Education, ECE test scores are one of the metrics used to rank schools and select the BE bonus recipients. Schools not eligible for taking the ECE test in 2015 (only 4% of all public secondary schools) compete for a smaller bonus based on other measures. We focus our analysis solely on public schools taking the ECE. As outlined in Table 1, a school's score for the BE is composed of several factors. The score gives 40% of weight to the average math and language grade of 8^{th} graders in the ECE standardized tests.[18] Additionally, 35% of weight is given to the entire school's intra-annual retention rate, that is, the proportion of enrolled students still in school at the end of the year. Although dropout rates are non-negligible, most of the dropping out takes places after the school year ends, making retention rates already extremely high before the program was implemented. The average retention rate in the public secondary schools was 99% in 2014, and only 7% of schools had retention rates below 95%. In practice, schools had very little leeway for improving their retention rates, and could thus not compete on the basis of this indicator.[19] An extra 5% of the school's score depends on whether the principal enrolls his students in the Ministry of Education's administrative system (*Sistema de Información de Apoyo a la Gestión de la Institución Educativa*, henceforth SIAGIE) in a timely manner, something which should not affect the incentives of teachers and thus the performance of their students. The remaining 20% of the score depends on an index of school management, composed of teacher attendance, management of school infrastructure, compliance with class hours, as well as measures of pedagogical practices and learning environment. The first three measures are collected by independent evaluators making visits to all public schools, whereas the last two are obtained from questionnaires handed out to 8^{th} grade students during the

[17]The ECE was first implemented in 2007 in 2^{nd} grade of primary school, and was extended in the following year to 4^{th} grade in schools with intercultural bilingual education. It was administered in 8^{th} grade for the first time in 2015, and will be extended to 4^{th} graders in all schools in 2016.

[18]In the primary school BE program, the score is also composed of the change in the average ECE scores from the previous year, to incentivize schools in the lower end of the distribution. Since 2015 was the first year the ECE was implemented in secondary schools, this could not be replicated.

[19]In conversations with the Ministry of Education, we were informed that due to operational limitations for calculating inter-annual retention rates, the retention rates employed in the BE program had to be calculated on an intra-annual basis.

ECE. All in all, around 80% of a school's score ultimately depends on the performance of 8^{th} grade students. In order to prevent teachers from encouraging absenteeism of low achieving students on the day of the ECE evaluation, schools not complying with a minimum rate of student participation are disqualified from taking part in the BE. In particular, ECE participation must be 80%, 90% and 95% in schools with only one, two or more than two 8^{th} grade classes.[20]

The timing of the BE is depicted in Figure 2. The school year in Peru starts in March, and ends in December. At the end of the 2014 school year, once the implementation of the ECE test in secondary schools was confirmed for the following year, the Minister of Education announced the possibility of extending the BE program to secondary schools as well.[21] The government resolution regulating the 2015 BE came out on the 25^{th} of July, almost four months before the 2015 ECE (carried out in November 17/18), and was accompanied by a diffusion campaign launched by the Ministry of Education informing schools about the BE program. In comparison to other studies, the time frame teachers had to react was relatively short. We elaborate on this issue using the results of our teacher survey in Section 7.6.

BE is set up as a collective incentive, such that the principal and every teacher in a school are rewarded if the school scores in the top 20%.[22] To ensure that schools competing against each other are comparable, they are separated into groups by school district, instruction time, and by whether they are urban or rural. There are 395 groups in total, with an average of 34.6 schools per group. Teachers in schools in the top 10% in their group get a bonus of 2000 soles (roughly 605 dollars), whereas those in schools in the top 10%-20% get paid 1500 soles (454 dollars). Every teacher in a winning school gets the exact same bonus, whereas the school principal gets a slightly larger payment (500 extra soles). Since the average teacher receives a monthly salary of 1469 soles, the bonus constitutes either 1 or 1.4 monthly salaries on average. Considering that 20% of schools receive the prize, the average value of the bonus is 24% of a monthly salary, a sizable figure as compared to other studies in the literature.[23,24]

[20] 91% of public secondary schools complied with this requirement, and the average school only had 1.4 students absent on the day of the exam.

[21] http://larepublica.pe/21-12-2014/jaime-saavedra-el-proceso-para-nombrar-a-8-mil-maestros-se-inicia-en-julio-del-2015 (last accessed August 16, 2016).

[22] Other papers studying collective teacher incentives are Lavy (2002) in Israel, Glewwe et al. (2010) in Kenya, Muralidharan and Sundararaman (2011) in India, Contreras and Rau (2012) in Chile, and Fryer (2013) and Goodman and Turner (2013) in the US.

[23] The average payment is 350 Soles (0.1x2000 + 0.1x1500), which constitutes 24% of the average teachers' monthly salary.

[24] In the teacher incentive program of Muralidharan and Sundararaman (2011), the average bonus was around 35% of

3 Estimation Strategy

We exploit the fact that a school's score for the BE largely depends on the performance of 8^{th} grade students in the math and language ECE test for estimating the causal effect of the teacher incentive on student learning. This feature of the BE results in schools having a much higher incentive to improve the learning of 8^{th} grade students as compared to students from other grades. With this notion in mind, we perform a difference-in-differences estimation comparing the change in achievement of 8^{th} grade public school students with that of 9^{th} grade students attending the same school.[25] In our preferred specification, we use a repeated cross-section of 8^{th} and 9^{th} grade students in public secondary schools eligible for the BE, and run the following regression:

$$Internal\ Grade_{ist} = \beta_0 + \beta_1\, 8^{th}\, Grade_{ist} + \beta_2\, 8^{th}\, Grade_{ist}\ \times\ Post_t + X_{ist}\,\delta + \gamma_t + \gamma_s + U_{ist}\ ,$$

where $Internal\ Grade_{ist}$ is the grade that student i in school s and year t obtained in a particular subject (i.e., the grade assigned to student i by his/her teacher at the end of the school year). We run separate regressions using math and language internal grades in our main specification, and also estimate this equation for the average internal grade in all subjects not evaluated in the ECE, to examine whether the BE impacted students' performance in other courses. $8^{th}\, Grade_{ist}$ is a dummy for whether student i from school s is an 8^{th} grader in year t, $Post_t$ is a dummy taking the value of one in the year 2015 and zero in 2013-2014, X_{ist} is a set of individual controls (gender, if Spanish is the student's native tongue, if the student was retained in the previous year, has a disability, and whether the parents are alive and living in the same household), and γ_t and γ_s are year and school fixed effects. We run regressions for the period 2013-2015, i.e., two years before the BE, and the year in which it took place. Our estimation thus compares students in 8^{th} and 9^{th} grade, within the same school, before and after the BE was introduced. Including school fixed effects allow us to restrict our comparison to students facing the same educational environment, but differing in their exposure to

a monthly salary; in the experiment run by Glewwe et al. (2010) in Kenya prizes were in-kind, and the average teacher got a bonus worth 12%-21% of a monthly salary. In the Israeli program studied by Lavy (2002) prizes of 10%-40% of the average teacher's monthly salary were awarded to approximately one third of participating teachers, whereas the prizes roughly represented 50% of a monthly salary and were awarded to two thirds of teachers in the New York experiment studied in Fryer (2013) and Goodman and Turner (2013). The incentive implemented in Chile and studied by Contreras and Rau (2012) awarded an average bonus of 10% of a monthly salary.

[25] We use 9^{th} grade as our comparison group and not 7^{th} grade, for example, because the program might have an impact on teachers in the latter grades, given that their students will be taking the ECE standardized test in 2016.

the BE.[26] U_{ist} are all the unobserved determinants of achievement for student i in school s and year t, such as ability, motivation, household income, and home environment, to name a few. We allow for our errors to be correlated within school by clustering our standard errors at the school level. We express grades as a z-score, standardizing them by subject and year, so that our coefficient of interest (β_2) can be interpreted as the standard deviation (SD) change in internal grades associated with the incentive. In the case of non-incentivized courses, we first calculate the z-score for each course, and then take the average. As a robustness check, we also standardize internal grades for each subject by school and year.

Unlike other studies on teacher pay-for-performance, our outcome variable is the grade assigned to students by their teachers at the end of the school year (what we refer to as internal grades), and not their standardized test results. Given that teachers' pay under the BE is tied to performance in the ECE, an identification strategy relying on standardized test scores as an outcome cannot disentangle whether improvements in students' performance are the consequence of increased student learning or the results of short-term strategies fostering high test scores (Neal, 2011). Having internal grades as our outcome has the advantage of capturing students' performance without directly influencing teachers' probability of obtaining the bonus. While it would still be interesting to study the impact of the BE on students' ECE test scores, we cannot do so because the ECE test was applied in secondary schools for the first time in 2015, and there is no group of students serving as an appropriate comparison. From the perspective of students, internal grades play a very important role, directly affecting whether they pass the school year, take summer remedial courses or are retained. Importantly for our identification, teachers' grading tactics should not be influenced by the BE incentive. Since the bonus from the BE is tied to ECE test results, and not internal grades, teachers' stakes in their students' internal grades are not directly modified by the incentive program.[27] Although it is mandatory for schools to report students' internal grades to the Ministry of Education, these grades have absolutely no bearing

[26] Given that 8^{th} grade students in private schools take the ECE but these institutions are not eligible for the BE, we could also run a differences-in-differences regression comparing the change of internal grades of 8^{th} grade students from public and private schools, similar to what Contreras and Rau (2012) do for the case of Chile. However, as shown in Appendix Table A.1, public school students were already improving relatively faster than their private school counterparts in the year prior to the BE (i.e., the *Public* $\times$ *2014* coefficient is statistically significant). Since there are other things that could be changing across the public-private spectrum in 2015 that we cannot control for, we discard this estimation strategy.

[27] Our teacher survey inquires, among other things, about whether teachers changed the difficulty of their classes in 2015 as a result of BE. As shown in Table 11, teachers were equally likey to report that they decreased the difficulty of their classes when teaching students from 8^{th} grade, as compared to students from other grades, and only 5 percentage points more likely to report that they increased the difficulty of their classes.

on whether the school obtains the bonus. As long as internal grades present enough variability, we would expect them to reflect changes in learning. We provide supporting evidence of the fact that within-schools, internal grades are correlated with standardized measures of learning, and show that grading on a curve is uncommon in Peruvian secondary schools in Section 6.2.

Our main identifying assumption is that in the absence of the teacher incentive, the performance of 8^{th} and 9^{th} grade students attending the same school would have evolved in an equivalent way between 2014 and 2015. A necessary condition for giving a causal interpretation to β_2 is that 8^{th} and 9^{th} grade students follow parallel trends before the implementation of the BE. An inspection of the raw means in Figure 1 shows that grades of 8^{th} and 9^{th} grade students appear to be on parallel trends in both math and language before the program was implemented. We provide formal evidence for the parallel trends assumption in Section 6.1.

Identifying a causal effect also requires that the performance of 9^{th} grade students, our comparison group, is unaffected or hardly affected by the teacher incentive program (i.e., that there are no spillovers). Importantly for our identification, schools do not have much room to compete on the basis of indicators other than the 8^{th} graders' standardized test scores, leading to a practically null correspondence between 9^{th} grade students' learning and a school's BE score. As explained in Section 2.2, around 80% of a school's score ultimately depends on the performance of 8^{th} grade students. This implies that, if any, a very small portion of the school's score could be improved if 9^{th} grade teachers exerted more effort. It is important to bear in mind, however, that since 83% of 8^{th} grade teachers also instruct 9^{th} grade, an increase in effort while teaching 8^{th} graders could potentially spill over to students in our comparison group and bias our estimates downwards. We show that this is not a concern by exploring the impact of the teacher incentive in schools with a low overlap between 8^{th} and 9^{th} grade teachers in Section 6.3. On the other hand, the fact that the probability of obtaining the bonus hinges largely on the performance of 8^{th} grade students might lead the school to redirect its resources towards these grades, negatively impacting the internal grades of students in our comparison group. We discuss this in further detail in Section 6.4, and show that this issue is not a concern in our setting.

4 Data and Descriptive Statistics

4.1 Administrative Data

Our empirical exercise relies on a rich administrative database collected by the Peruvian Ministry of Education in 2013-2015, derived from its SIAGIE system. Coverage is basically universal, reaching 99.7% of public schools. Schools must enroll their students into the SIAGIE system at the start of the school year, and input the final grades of their entire student body once the academic year concludes. Grading is done on a 0-20 scale, and students need to obtain at least 11 to pass a given subject. Besides students' grades, this database also has information on characteristics such as age, gender, native tongue, parents' education, if they live with their parents, etc. Student identifiers permit tracking individual students across years. The SIAGIE also contains information on the grade and classroom that student are assigned to, the teachers who teach each grade and group, and some basic teacher characteristics such as age and gender. In 2015, there were 8,654 public secondary schools in Peru, of which 8,092 were eligible for participating in the ECE. Schools must have at least five 8^{th} grade students in order to be eligible for taking the test. Our SIAGIE database covers 8,059 of these schools.

Table 2 presents some characteristics of the 8^{th} and 9^{th} grade students attending public secondary schools in 2013-2015. We observe that the mean final grade in math is 12.27 and 12.32 (out of 20) in 8^{th} and 9^{th} grades, and 84% and 85% of students pass this course. Students perform slightly better in language, where the mean final grade for the 8^{th} and 9^{th} grade students is 12.67 and 12.64, and 89% and 90% of them pass the course. Mean grades in other courses exceed those of math and language by almost one point, and almost all (93% and 94%) students pass these courses. Half of the students are male, almost all of them are natives, and 83-84% of them have Spanish as a native tongue. Only 6% and 4% of 8^{th} and 9^{th} graders were retained in the same grade the year before. Although it is not necessary for our identification that 8^{th} and 9^{th} grade students are balanced in terms of observables, they do appear to be very similar. In addition, Table 2 shows some characteristics of the 8,059 public secondary schools in our sample. Less than half (40%) of the public secondary schools are located in rural areas. Each school has, on average, two classes per grade, and there are around 19 students per teacher in the average class. We also observe that each school has, on average, roughly 11 teachers teaching 8^{th} and 9^{th} grade, with 83% of teachers in 8^{th} (9^{th}) grade also teaching 9^{th} (8^{th}) grade.

Teachers are almost 42 years old on average, and 60% of them are male.

4.2 Survey Data

We complement our main empirical analysis with the results of an online survey we conducted with the assistance of the Ministry of Education. According to our teacher database, there were 123,669 public secondary school teachers in 2015. The Ministry of Education has the email address of 36,283 of them (30%), all of which received a survey email from the Ministry in October 2016, a few weeks before the winners of BE were announced. As in the past editions of BE in primary schools, the bonus winners were announced at the end of the following school year (in November 2016). Teachers were asked what grades and subjects they taught in 2015, their knowledge about the BE and its rules at that time, and their opinion about the size of the bonus. We also inquired about changes in their pedagogical practices while teaching students from different grades, and about administrative changes in the school they were working for in 2015. Finally, we tried to elucidate teachers' perception about their school's ranking and its probability of winning, and asked teachers for their opinion about students' motivation in the standardized test tied to the BE.

The survey was anonymous, and teachers were told that its purpose was to collect information about teachers perceptions and opinions about the BE program. Since the survey was framed in the context of BE, and sent by the Ministry of Education, respondents might be subject to social desirability bias (i.e., over-reporting of good behavior associated with the objectives of BE). To try and maximize the response rate, and due to restrictions imposed by the Ministry, we did not ask questions about teacher characteristics or identify the school they worked for, and thus we cannot compare survey respondents to non-respondents. We received a response from 3,406 teachers (9.4% response rate), roughly 2.8% of all public secondary school teachers. Given the potential bias in teachers' responses and our selected sample, the results from this survey must be taken with caution.

5 Results

The teacher incentive program had no effect on 8^{th} grade students' math and language internal grades, as shown in columns (1) and (4) of Table 3. Our coefficient of interest (the interaction of the 8^{th} *Grade* and *Post* dummies) is robust to the inclusion of school fixed effects (columns 2 and 5) and individual controls (columns 3 and 6), with the latter being our preferred specification.[28,29] Our coefficients are precisely estimated zeros, allowing us to reject positive effects larger than 0.008 SD in math, and 0.017 in language, well below the treatment effects found in the existing literature. In the teacher incentive program studied by Muralidharan and Sundararaman (2011) in India, average math and language test scores increased by 0.15 SD after one year, whereas Contreras and Rau (2012) find that a teacher incentive program in Chile had positive and large effects on language and math test scores of 0.14-0.25 SD. While the incentive scheme evaluated by Glewwe et al. (2010) in Kenya led to a 0.14 SD increase in test scores in tests linked to the incentive, the authors found no impact on the outcome of non-incentivized evaluations, consistent with our findings.

Since there are 395 distinct groups in which schools compete for the BE bonus, i.e., 395 different tournaments, we also evaluate the average effect of the teacher incentive in every competition. Figure 3 displays the 8^{th} *Grade* x *Post* coefficients (and its 95% confidence interval) for math and language in each of these 395 tournaments. In the vast majority of these groups, the teacher incentive had a zero average effect on student achievement. The coefficients for math and language are positive and statistically significant at the 5% level in only 4% and 6% of the BE groups,[30] providing further evidence of the BE's null average effect on student achievement.

As in most comparable studies, teacher bonuses under the BE are tied to students' performance in just two subjects (math and language). However, teacher incentives might also have an impact on student learning in other courses. The sign of this impact is theoretically unclear. On the one hand, schools could be tempted to devote more resources towards math and language at the expense of other subjects (e.g., augmenting instruction time), negatively impacting learning in the latter. On the other hand, 8^{th} grade teachers in all subjects, not just math and language, might exert more effort knowing that their school's score largely rests on the performance of these students. Additionally, due to complementarities, if learning were higher in math and language, student achievement in incentivized subjects might increase indirectly (Muralidharan and Sundararaman, 2011).[31] We do find a positive and small but significant effect of 0.011 SD on grades in non-incentivized courses, as shown in Table 4.

[28]In our baseline regressions we standardize internal grades by subject-year, but the results are quantitatively similar if we standardizing each subject by school and year.

[29]Parents' education is missing for 12% of students, so we do not control for this in our baseline regressions. However, attrition is not differential across grades, and our results are robust to controlling for this.

[30]Furthermore, in only 6 out of the 395 BE groups this holds simultaneously for math and language.

[31]Unlike studies carried out in primary school, math and language teachers are not responsible for teaching other subjects in secondary school. Thus, if there were any positive spillovers to other courses, they would be indirect.

Appendix Table A.2 breaks the results down by each of the nine non-incentivized courses; we observe positive effects ranging between 0.014 SD and 0.017 SD in three cases (social studies, human relations, and religion). Although significant, the observed effect is very small, and well below the spillover effects found in other papers.[32] Furthermore, these results should be taken with caution because, as further discussed in Section 6.1, there is a divergence in the trend in non-incentivized courses in the year before the program was implemented.

5.1 Heterogeneous Effects

In a tournament such as the BE, if teachers are risk neutral, have symmetric information, and if students in all schools have the same ability (i..e, if all schools have the same ex-ante probability of winning), all teachers should exert the same effort as a result of the incentive, and who gets awarded the bonus should be random (Lazear and Rosen, 1981). However, if schools differ in their probability of winning, the incentive might not have the same power across the board. For example, teachers in schools in which students' pre-program levels of achievement are very far from the top 20% will be discouraged from exerting extra effort. Alternatively, schools which are almost guaranteed to win might not be motivated by the pay-for-performance program. This concern is partly mitigated in our setting by the fact that schools are grouped according to characteristics which are likely correlated with their students' performance, such as instruction time, by whether they are urban or rural and by their school district. However, schools within each group might still be heterogeneous, possibly affecting the program's reach. This notion is brought forward in the Chilean study of Contreras and Rau (2012), where the authors find that the teacher incentive only had a positive impact on schools above the 65^{th} percentile in the distribution of pre-program score (the program awarded a bonus to schools in the top 25% within their group). The fact that the ECE was implemented in secondary schools for the first time in 2015 provides a limitation for studying this, since it does not allow us (or the schools themselves) to determine a school's pre-tournament probability of winning. As a second best, we proxy a school's likelihood of winning using its relative ranking within its BE group in terms of the socioeconomic status (SES) of students. We construct a measure of 8^{th} grade students' average SES considering whether their first language is Spanish, and whether their parents have more than

[32] Muralidharan and Sundararaman (2011) find that teacher incentives targeted towards math and language standardized tests had an effect of 0.11 and 0.14 SD in science and social studies after only one year, an effect 10 to 13 times larger than the one we find.

a primary school degree.[33] As shown in columns (1)-(3) of Table 5, there is no differential effect on student achievement in math and language for schools above the 55^{th}, 65^{th} or 75^{th} percentile in terms of their 8^{th} graders' SES, as compared to schools within the same BE group whose students have a lower SES. The former schools are arguably those that had a good chance of winning the bonus. If we treat schools in the top 5% of their BE group (the sure winners) differently, we still observe no differential effect for schools in the 55^{th}-95^{th}, 65^{th}-95^{th}, and 75^{th}-95^{th} percentiles.[34] Having said this, it is highly likely that schools did not know their relative standing in their BE group, and could thus not anticipate the likelihood of winning. We discuss this in Section 7.5.

From a theoretical perspective, the strength of the incentive might be decreasing in the number of 8^{th} grade teachers and/or students, since the marginal impact of a teachers' effort on its school's score decreases when there are more teachers and students reached by the incentive, and teachers' ability to monitor each other also diminishes. For instance, Imberman and Lovenheim (2015) find that the effect of a group-based teacher incentive program in Houston is much stronger when teachers are responsible for teaching a higher share of students. Since our teacher database does not have information on the subject that each teacher is responsible for, we do not know how many incentivized teachers each school has; as a second best, we use the number of 8^{th} grade classes in 2015 as a proxy. We do not find any significant interaction of the BE incentive with enrollment or number of groups per grade, as seen in columns 6 and 7 of Table 6. Finally, we do not find any effects by whether the school ir urban or rural, as shown in column 8. As with any heterogeneity analysis, it is important to take the results with caution, since characteristics such as enrollment and urbanicity are not randomly assigned, and could be proxying for something else. Ideally, we would also be able to test for heterogeneous effects across teacher characteristics. Unfortunately, although we know who the teachers are for each class, we do not know which of the teachers teach math and language.

Following other papers in the literature, we also test for heterogeneous effects across gender, by whether students' first language is Spanish, and by their parents' educational attainment. The latter variable is an index from 0 to 2, taking a value of 0 if both parents have a primary school degree or less, 1 if one parent has more than primary schooling, and 2 if both do. Parents' education and students'

[33]For each 8^{th} grader in 2015, we add three dummy variables: whether his first language is Spanish, and dummies for whether his mother and father have more than a primary school education. We then calculate the average index for each school, and stratify schools according to their ranking within their BE group.

[34]We perform the same exercise in Appendix Table A.4 stratifying schools according to an index measuring the quality of their infrastructure, another proxy of their probability of winning, and find similar null results.

native tongue are proxies for socioeconomic status in Peru. As displayed in column 1 of Table 6, and consistent with the finding in Muralidharan and Sundararaman (2011) and Behrman et al. (2015), we do not find any heterogeneity by gender. Neither do we find heterogeneous effects by socioeconomic status, proxied by native language and parents' education (columns 2 and 3). The literature is mixed on this particular issue, since Muralidharan and Sundararaman (2011) observe that students from more affluent families have a stronger response to the teacher incentive program, whereas Lavy (2002) finds that it is students with poor socioeconomic backgrounds that benefit more from it. However, the program evaluated in the latter was designed so as to encourage teachers to focus on weak students.

Considering that teachers might focus on certain students, and student responsiveness might vary according to prior achievement, we also test for heterogeneity across measures of students' past performance, namely whether the student was retained in the previous year and by the student's lagged internal grade in the same subject (standardized by school, grade and year).[35] Pay-for-performance programs in which bonus payments depend on whether students attain a certain threshold, such as passing an exam, create incentives for teachers to focus on students close to this cutoff (e.g. Lavy, 2009 and Neal and Schanzenbach, 2010). On the contrary, if obtaining the bonus depends on the average score, such as in the BE program under analysis, teachers will find it optimal to target students most responsive to any increased teacher effort. If the function mapping teacher effort into test score gains is concave (convex) in past performance, teachers would react by focusing more intensely on the weaker (stronger) students (Muralidharan and Sundararaman, 2011). However, as shown in columns 4 and 5, we do not find any heterogeneity according to students' past performance.[36] These results are consistent with the findings of Behrman et al. (2015).

6 Testing the Validity of the Identification Strategy

This section provides further evidence on the validity of our difference-in-differences estimation strategy. We provide formal evidence in support of the parallel trends assumption, and demonstrate that internal grades are broadly correlated with ECE test scores, and vary considerably within schools. Furthermore, we corroborate that our null effects are not driven by positive spillovers to our com-

[35] Lagged grades are only available for students in 2014 and 2015, since our database only has student identifiers which can be linked across years starting 2013. Importantly, if we restrict our sample to this period, results on average treatment effects do not change.

[36] We also perform this estimation by grouping students into quintiles and terciles of the distribution of lagged grades in their same school, grade and year. The results are unchanged, as reported in Table A.3.

parison school, and show that schools did not change the way in which they assigned teachers across grades as a result of the teacher incentive program.

6.1 Parallel Trends

To test whether there is a divergence in the trends of 8^{th} and 9^{th} grade students in 2014, we add an interaction between the 8^{th} grade dummy and an indicator for 2014 to our baseline specification. Reassuringly, the coefficients for the pre-treatment difference-in-differences are precisely estimated zeroes for both math and language, as shown in Table 7. In the case of non-incentivized courses, however, there is a relative increase in 8^{th} graders' internal grades in 2014. Although the magnitude of this change is small (0.010 SD), it is similar in magnitude to the estimated impacts for 2015. Hence, the results using non-incentivized courses as an outcome should be taken with caution.

6.2 Internal Grades Reflect Learning

Unlike other studies on teacher pay-for-performance, we measure learning using students' internal grades instead of their standardized test results.[37] As discussed in Section 3, internal grades have the advantage of capturing student achievement without directly influencing teachers' probability of obtaining the bonus. However, internal grades are subjectively assigned by teachers, and are not awarded using a uniform criterion as standardized tests are. Since each school might have its own grading standards, making differences in internal grades not necessarily reflective of differences in learning across schools, we restrict our comparison to students from the same school to control for school-specific grading standards.[38] What is crucial for identifying a causal effect is that internal grades capture changes in learning across different grades within the same school. That is, if 8^{th} grade students in a particular school are learning more as a result of the teacher incentive, the relative internal grades of 8^{th} graders in that school should rise. We face two potential threats in this regard. Firstly, teachers might not award internal grades in a systematic way. This doubt is raised by the findings of a few papers comparing grading standards in blind versus non-blind examinations. While some studies find evidence of discrimination in grading based on students' gender (Lavy, 2008), ethnicity (Botelho

[37] In a recent study, Chong et al. (2016) also use internal grades to measure student achievement in rural Peru.

[38] Although it would be preferable to include teacher fixed effects to control for teachers' grading standards, we only know the grades and classes teachers are assigned to, but not the subject that they teach. We cannot identify who the teacher handing out the grades for each subject is, and therefore cannot include teacher fixed effects in our estimation. However, since teachers are not systematically changing across 8^{th} and 9^{th} grades, as shown in Section 6.4 below, unobserved teacher characteristics are unlikely to bias our estimates.

et al., 2015; Burgess and Greaves, 2013), and caste (Hanna and Linden, 2012), others find no such disparities (Newstead and Dennis, 1990; Baird, 1998; Van Ewijk, 2011). Consistent with the latter, we show that student characteristics correlate with internal grades and with standardized test scores in a consistent manner within the schools in our sample, alleviating this concern. A second threat to our identification is that if teachers grade on a relative basis (e.g., the worst 10% always fails, or the top 10% always gets the highest grade), we might not be able to detect overall changes in student learning using internal grades. It turns out, however, that there is substantial variation in the distribution of grades across classes and years in the same school.

Considering that our identification requires that internal grades reflect within-school differences in learning, standardized test scores and internal grades should broadly follow the same patterns when comparing students from the same school. Unfortunately, ECE test scores are disclosed at the school level, meaning that for every secondary school taking the ECE in 2015, we only observe the mean score in math and language, as well as the fraction of 8^{th} graders with very low, low, medium and high performance. Given our data limitations, we examine the cross-sectional correlation between average ECE scores and the average internal grades of 8^{th} graders in 2015, for the public secondary schools in our sample.[39] We also explore the correlation between the fraction of students who fail math and language according to their internal grades, and the fraction of low performing students in the ECE. To facilitate the interpretation of the coefficients, we express average internal grades and average ECE scores as a z-score, and control for school district fixed effects, school characteristics and the average characteristics of students from each school. As shown in Appendix Table A.5, average ECE scores and internal grades are positively and significantly correlated, although their correspondence is relatively weak. In particular, a 1 standard deviation increase in average math (language) internal grades is associated with an increase in average ECE scores of 0.116 (0.103) SD. Moreover, a 1 percentage point increase in the share of students failing math (language) according to their internal grades corresponds to a 0.071 (0.091) percentage point rise in the proportion of students with the lowest attainment in the ECE.[40,41]

[39] Our sample for this analysis (8,010 schools) is slightly smaller than our baseline sample of 8,059 schools because a few schools with were eligible to take the ECE (and were thus eligible to participate in the BE) ended up not taking the test, or were faced with problems during its implementation.

[40] In 2015, 55% and 43% of 8^{th} graders in the average school were ranked in the lowest category according to their ECE scores, whereas the average school only had 13% and 9% of their 8^{th} graders failing math and language, respectively. These two categorizations are only broadly comparable, and these results must thus be taken with caution.

[41] We also examine whether school and average student characteristics explain internal and ECE grades in a similar manner, by separately regressing schools' 2015 average ECE and internal grades against a series of controls. As displayed

Having said this, it is hard to establish whether internal grades reflect learning by just comparing the aggregate cross-sectional correlation of these and ECE grades. For one, internal grades might capture a related but different dimension of learning than standardized test scores. Additionally, since internal grades are likely to depend on school grading standards, it is unclear that they can be compared across schools.[42] While the disclosure of ECE test scores does not allow us to identify students' individual performance, the Peruvian Ministry of Education provides an anonymized database with individual ECE test scores, gender, an index of socioeconomic status (constructed using parents' education, and household assets and characteristics), and anonymized school identifiers. As shown in Panel A of Table A.7, students are more likely to obtain a higher ECE test score in math and language if they are male and have a high socioeconomic status, as compared to other students from the same school. An analogous regression with 8^{th} grade students' individual internal grades as the dependent variable (Panel B of Table A.7) shows that the within-school correlation between student achievement and gender and socioeconomic status is qualitatively similar. Despite the fact that internal grades and standardized test scores are prone to measure learning differently, and that students have different stakes in each of these outcomes, these two measures seem to relate in a consistent manner when comparing students from the same school.

Having established that internal grades are correlated with standardized test scores, we now provide evidence of the fact that grading on a curve is uncommon in Peruvian secondary schools. If teachers were assigning grades on a relative basis, we would expect two different classes in the same school, grade and year to have a very similar grade distribution. Our database on teachers shows that on average, 8^{th} grade teachers from schools with only two classes teach in 92% of them, meaning that the teachers handing out the grades are practically the same across classes. We restrict our sample to 8^{th} graders in schools with just two 8^{th} grade groups in 2014 (accounting for 17% of our schools), and test whether math and language internal grades have a different mean and standard deviation across both classes belonging to the same school. With a significance level of 10%, in 23% and 32% of cases we reject the null hypothesis of equal means across both groups in math and language, respectively.

in Appendix Table A.6, the same broad patterns hold for both types of grades in math and language. Schools in which a high proportion of students have parents with more than a primary school degree do better as reflected by both ECE and internal grades. The same holds for schools with longer school days, and schools in which a high proportion of students have Spanish as their first language, Furthermore, schools in which a higher share of 8^{th} graders were retained the year before do worse according to both measures.

[42] If good schools set harsh grading standards, and low quality schools are lenient in their grading, for example, differences in the average internal grades of these two types of schools will not convey any information on their differences in student achievement.

The average difference in means across groups is 0.66 and 0.77 in math and language, roughly one third of a standard deviation. An F-test for the equality of variances shows that in 23% and 21% of our schools, we can reject the null hypothesis that the distribution of math and language grades has the same standard deviation.[43] The difference in means and standard deviations and their corresponding p-values are depicted in Figure 4. All in all, this evidence points to the fact that grading on a curve is not the norm in Peruvian high schools.

6.3 No Spillovers to 9^{th} Grade Students

A possible explanation for our null effects is that while BE has an impact, it improves teacher behavior overall, and does not impact teaching to 8^{th} graders differently. As we explain in Section 2.2, in practice around 80% of a school's score depends on the performance of 8^{th} grade students in the ECE standardized tests. Thus, a very small portion of the school's score could be improved if 9^{th} grade teachers exerted more effort. However, since 83% of 8^{th} grade teachers also instruct 9^{th} grade, any increase in effort while teaching 8^{th} graders could spill over to students in our comparison group, biasing our estimation downwards. Alleviating this concern, we find that the effects are also null in schools in which a low share of 8^{th} grade teachers also instructs 9^{th} graders (columns 1, 2, 4 and 5 of Table 8).[44] If anything, there is a significant (though very small) positive effect in math grades in schools in which most teachers instruct both 8^{th} and 9^{th} graders (column 3).

On the other hand, the fact that the probability of obtaining the bonus hinges on the performance of 8^{th} grade students might lead teachers to redirect their efforts to these grades, negatively impacting the internal grades of students in our comparison group and creating an upward bias in our estimations. However, our null results show that this concern is also irrelevant in our context.

6.4 No Differential Change in Teacher Composition

The incentives introduced by BE could have led schools to change the assignment of teachers across 8^{th} and 9^{th} grade in 2015. If schools assigned the best teachers to 8^{th} grade in 2015, for instance, our results would be upward biased. Since the program was only announced during the middle of the school year, as illustrated in Figure 2, it must have been hard for schools to shift teachers around.

[43] The average difference in standard deviations is 0.44 in math and 0.41 in language.

[44] Even though what matters is the overlap of math and language teachers, we do not have information on the subjects taught by each teacher, and are thus restricted to perform this analysis using the average overlap of all teachers across 8^{th} and 9^{th} grade.

It is therefore unlikely that this is a big concern in the first year of the program. Given that we cannot identify which teachers are responsible for instructing math and language, and do not have a measure of teacher quality, it is hard to test if BE brought about changes in the average quality of teachers across grades. However, we observe the school, grades and classes to which teachers are assigned in 2013-2015, and have some observable teacher characteristics which might be correlated with their performance. We use this information to test for differential changes in 2015 in the average characteristics of 8^{th} and 9^{th} grade teachers from the same school. As shown in Table 9, we do not find any differential change in the average age and gender of teachers in 2015. Neither do we observe a significant change in the proportion of teachers who are new to the school or new to that particular grade and school, or in the average number of courses taught by teachers in that grade. We do observe a significant decrease in the average number of secondary schools in which 8^{th} grade teachers are working. Although this might mean that 8^{th} grade teachers were less time constrained in 2015, this only represents a 0.3% drop from the mean. Consequently, there is no strong evidence of changes in teacher composition across 8^{th} and 9^{th} grade classes belonging to the same school in 2015.

7 Why Didn't Student Learning Increase?

Having established that student learning did not increase as a result of the teacher incentive program, this section discusses and provides suggestive evidence on a series of potential explanations for why the program had a null effect.

7.1 Teachers Did Not Know About the Program or Did Not Understand it

An explanation for the null effects we find is that schools simply did not hear about the BE, or did not understand the formula by which scores were calculated. We argue that this is at best a partial explanation. In 2015, along with the launch of the 2015 edition of the BE program, the Ministry of Education headed a diffusion campaign, making it likely that secondary school teachers were informed about the program. Furthermore, the fact that the principal and every teacher get paid if their school wins generates strong incentives for people working in the same institution to inform each other about the BE. In our teacher survey, 64% of those who taught math or language in 8^{th} grade in 2015 reported that they knew about the program's existence during the 2015 academic year. When asked about how they heard about BE, 57% answered that they found out through the Ministry of Education or the school district authorities, 30% answered that they got the information from the news, and 35% from the school principal or other coworkers (they could select more than one option).

Although we only evaluate the effect of BE in its first year, the system by which schools were scored under the BE was not overly complex.[45] It should have been relatively clear from a teacher's perspective that the main component of his/her school's score is the average performance of 8^{th} graders in the ECE standardized tests. This stems from the fact that performance in the ECE test was the main component of schools' scores in the two previous rounds of the BE in primary schools. The BE program had already been going on for two editions in every public primary school in the country, and the experience of primary schools with the BE was salient in the national news.[46] This is broadly confirmed by our survey, in which 64% of math or language 8^{th} grade teachers who knew about BE in 2015 answered that ECE test scores were the most important or second most important component of the BE score.

Almost half of the schools in our sample share the building with a primary school that participated in the BE before, and 13% of them operate in the same building as primary school BE winner. Even though the salience of the secondary school BE was probably higher in these cases, we do not find any effects on math and language test scores in either of these groups of schools, as shown in columns (3), (4), (8) and (9) of Table 10. Thus, it is unlikely that the BE had no impact because of schools' lack of awareness of its existence or its rules.

7.2 The Incentive Was Too Small

The prize that teachers could receive under the BE is in the range of bonuses granted in other studies finding positive effects. As described in Section 2, the BE bonus corresponds to either 1 or 1.4 monthly salaries of the average teacher, and was awarded to teachers in 20% of schools. The average bonus represents approximately 24% of a monthly salary, making this incentive sizable in comparison to that of other studies, in which the average value of the prize ranges between 3% and 35% of a monthly salary.[47] In the subsample of 8^{th} grade math or language teachers who responded our survey, 42%

[45] Other studies on teaching incentives with similar formulas for assigning the bonus (Lavy, 2002 and Contreras and Rau, 2012) find positive and significant effects on student learning.

[46] For instance, http://larepublica.pe/23-10-2014/maestros-tendran-bono-de-hasta-3-mil-soles-por-buen-desempeno and http://www.andina.com.pe/agencia/noticia-bono-hasta-s-3000-buen-desempeno-docente-se-pagara-noviembre-528482.aspx

[47] The average bonus obtained by Indian teachers in Muralidharan and Sundararaman (2011) is around 35% of a monthly salary, whereas bonuses in the experiment run by Glewwe et al. (2010) in Kenya have an average value ranging between 12% and 21% of a teachers' monthly wage. The incentive implemented in Chile and studied by Contreras and

correctly identified the bonus amount or thought that it was larger, 20% did not know the exact bonus amount, 2% thought that it was smaller, and 36% did not know about the BE in 2015. However, when we asked their opinion on the size of the prize, only 30% of those who knew about the program thought that the prize was adequate or large. This may have to do with the fact that the survey was coming from the Ministry of Education, and many teachers took this as an opportunity to complain about their low salaries.[48]

If the bonus were not large enough to incentivize the average teacher, we would perhaps find a positive effect in schools in which teachers' pay is relatively low. However, as shown in columns (1) and (6) of Table 10, we do not find any heterogeneity by teachers' average salary in 2015.[49] Although we cannot exclude that the incentive scheme would have worked with a larger bonus, there is no evidence that the size of the incentive is the reason why the program had no distinguishable effect on students' math and language grades.

7.3 Group Incentives Do Not Work

When incentives are collective, the mapping of a teachers' actions on his/her probability of obtaining the bonus is weaker when the number of teachers reached by the incentive is larger, raising the likelihood of free-riding (Holmstrom, 1982), and thus lowering the incentive's power in promoting higher teacher effort. While collective incentives have the potential of inducing higher cooperation and monitoring among teachers (Kandel and Lazear, 1992; Kandori, 1992), this might be harder to achieve when the number of incentivized teachers is very large. Although we do not know the fraction of 8^{th} grade students that each math and language teacher instructs (we do not have information on the subject taught by teachers), we do know the number of 8^{th} grade classes that each school has in 2015. In 2013-2015, the average secondary school in our sample had only two groups of 8^{th} graders. Since there is at most one math and language teacher per group, the average school has no more than four incentivized (i.e., math and language) teachers, a figure comparable to the number of incentivized teachers in other papers in the literature finding positive effects when teacher incentives are collective.

Rau (2012) paid teachers 10% of a monthly salary on average. Finally, the Israeli program studied by Lavy (2002) awards prizes of 10%-40% of an average teacher's monthly salary to approximately one third of participating teachers.

[48] In the open-ended part of this question, many teachers answered that their salaries are insufficient. Furthermore, quite a few teachers answered the survey email with complaints about their working conditions.

[49] We calculate the average salary of teachers in every secondary school from the number of contract teachers and civil servant teachers in each pay scale, as reported in the 2015 school census. Since the school census does not provide disaggregated data by grade, we take the school average.

The average school in Muralidharan and Sundararaman (2011) and Glewwe et al. (2010) has three and six incentivized teachers, for example.

As shown in column (7) of Table 6, we don't find any differential effects by the number of 8^{th} grade groups. If we break the results down even more, as shown in column (2) of Table 10, we do find that the BE had a small but significant positive effect in the math grades of students in schools with only one class per grade (accounting for 21% of students in 61% of schools). More specifically, the teacher incentive increases math grades by 0.019 SD,[50] although these effects are much smaller than those found in the other studies in the literature. Thus, the fact that the incentive faced by teachers under the BE is collective does not seem to be one of the main reasons why the program had no effect, although it might have some bite.

7.4 Teachers Only Focused on Improving Standardized Test Scores

As discussed in Section 3, teacher incentive programs might not result in higher learning if teachers focus their efforts on short-term strategies aimed solely at increasing standardized test scores. Teachers might have reacted to the incentive by targeting topics likely to appear in the ECE, coaching students on test-taking strategies, or even cheating. Since 2015 was the first year in which students took the ECE, and there is no appropriate control group (every public school in which 8^{th} graders participated in the ECE is also eligible for the BE), we cannot identify whether ECE test scores increased as a result of the teacher incentive program. Thus, we cannot initially rule out this hypothesis. However, there are reasons why we believe that teachers could not engage in this type of behavior. Firstly, independent officials, trained and working directly for the Peruvian National Statistics Institute were in charge of the implementation of the ECE. Teachers were not allowed to be in the room at any moment during the exam and were not responsible for its correction. Thus, it is very unlikely that schools could cheat.[51] Secondly, because the ECE exam is designed to capture a wide range of skills,[52] teachers could hardly influence this outcome by narrowing their instructional focus on certain topics. Thirdly, due to the fact that the ECE was implemented for the first time in secondary schools in 2015, secondary school teachers did not have previous experience with this type of standardized tests and,

[50] The sum of the 8^{th} $Grade \times Post$ and 8^{th} $Grade \times Post \times One\ Class$ coefficients yields a total effect of 0.019 SD, with a p-value of 0.008.

[51] Since students had no personal stake in this exam, there were no incentive to cheat on their part.

[52] Details on the design of the ECE are reported by the Ministry of Education in *Reporte Técnico de la Evaluación Censal de Estudiantes (ECE 2015)*, available at http://umc.minedu.gob.pe/wp-content/uploads/2016/07/Reporte-Tecnico-ECE-2015.pdf.

consequently, could hardly predict the content or the specific format of the exam. As the content of the standardized exam was not predictable, coaching or narrow teaching are less of a concern in this setting (Neal, 2011).

Having said this, our online survey inquired about whether teachers changed their pedagogical practices in 2015 as a result of BE, and separately asks about their pedagogical changes while teaching 8^{th} grade as opposed to all other grades. Table 11 reports the results of this question for all math and language teachers taking the survey who reported that they knew about BE in 2015 (those who did not know where not asked this question). These results must be taken with caution, since it is probable that there was some bias in reporting given the framing of the survey in terms of the BE program.[53] As can be seen in Panel A, 8^{th} grade teachers are 5 percentage points more likely to report that they improved their attendance, and 10 percentage points more likely to report that they gave their students more homework, evaluated them more often and/or gave extra tutoring sessions, as compared to math and language teachers from other grades. There are statistically significant differences as well in how often they report that they paid attention to the weakest students (5 percentage points), increased the difficulty of their classes (6 points), and increased the frequency of multiple choice examinations (9 percentage points). The same patterns hold when we restrict the analysis to teachers that taught math or language in 8^{th} grade and other grades, as seen in Panel B. While some of these self-reported differences in teacher behavior are consistent with teaching-to-the-test (e.g., increasing the frequency of multiple choice evaluations), if teachers were in fact improving their attendance or paying more attention to the weakest students, student achievement in terms of internal grades should have increased, and it did not.

7.5 Teachers Were Unfamiliar with the Standardized Test and Students Had No Stakes in it

Given that 2015 was the first year in which the ECE test was implemented in secondary schools, teachers might have been uncertain about the function mapping their effort into students' ECE test scores. The connection between teachers' effort and their expected benefit might have therefore been

[53] In the study of Glewwe et al. (2010), for example, the survey to teachers was also framed as soliciting feedback on the incentive program; teachers in the treatment group were more likely to report having increased the number of homework assignments, whereas student reports suggest no such differences. In Behrman et al. (2015), teachers were also more likely to report that they spent more hours preparing their students for the test, although the incentive had no impact on student outcomes.

diluted, making the incentive insufficient for prompting teachers into exerting more effort (Fryer, 2013).[54] Even if teachers knew how to equip their students with the skills needed to obtain high ECE scores, they might have encountered difficulties in passing on the incentive to their students. Since ECE tests have no impact whatsoever on students' grades, and the Ministry of Education only reports school averages (and not individual test scores) to schools, teachers, parents and even students, the latter might have little or no incentive to put effort in these tests.[55] Teachers might have anticipated that their actions would only marginally impact their students' ECE scores, and thus might have been discouraged from exerting more effort. The results from the experimental study implemented by Behrman et al. (2015) in Mexico provide suggestive evidence on the possibility that incentivizing teachers on their students' performance might not be effective unless students have a stake as well.[56] This hypothesis is partially supported by our online survey to teachers. When asked whether they thought students put any effort when taking the ECE test, 37% of survey respondents who taught math or language in 8^{th} grade answered that they did not. We inquired about the reasons for why students do not put any effort while taking the ECE, and teachers replied that this was due to the fact that ECE test scores do not affect their final grade (51%), because students are unmotivated (47%), the test is too long (10%) or too difficult (8%), and students are not familiar with these types of evaluations (11%).

Since the ECE was implemented in secondary schools for the first time in 2015, schools might not have known the relative standing of their students in comparison to those from competing schools. Teachers might have been unable to infer the level of effort needed for their school to win the bonus, thus lowering the power of the incentive and ultimately discouraging them from putting in more effort.

[54] A series of experimental studies in rural India suggest that teachers' knowledge and incentives might be complementary inputs in the education production function. While Muralidharan and Sundararaman (2010) show that giving teachers feedback on their students' past performance and detailed information on how to improve students' learning in low stakes tests has no effect on tests scores, students' test scores increased when this informational treatment was paired with a monetary incentive to teachers based on the performance of their students (the treatment from Muralidharan and Sundararaman, 2011). Since there is no treatment arm with monetary incentives but no information, it is hard to disentangle whether this effect is simply due to the monetary incentives, or whether the latter are only effective when teachers are given enough information on how to influence student learning.

[55] The findings of Tran and Zeckhauser (2012) and Azmat and Iriberri (2010) are consistent with the notion that not informing students about their achievement in the ECE might keep them from applying themselves while taking the test. Both studies find that providing students with relative performance feedback enhances their performance, even if they are not rewarded for it.

[56] The evidence provided by Behrman et al. (2015) on this point is only suggestive because, as compared to the treatment in which only teachers were incentivized, the potential reward for teachers and students was larger in the treatment arm in which both were incentivized. It is therefore hard to tease out if this incremental effect is coming the existence of complementarities between teachers' and students' effort, or from the fact that the monetary incentives were larger.

Since schools participating in the BE compete against other comparable schools within their district, they might have some prior about how their students compare to those of the competing schools, especially in BE groups with few schools. As shown in columns (5) and (10) of Table 10, there is a small but positive effect (0.012 and 0.023 SD in math and language) on student learning in schools in BE groups smaller than 27, the median group size. Although group size is probably an inaccurate proxy for knowledge about the probability of winning, this suggests that it might be important for schools to know how much effort they need to exert for the program to be effective.[57]

7.6 Teachers Did Not Have Enough Time to React

Finally, schools might not have had enough time to increase their students' learning in a meaningful way. As explained in Section 2, the Minister of Education mentioned the possibility of extending BE to secondary schools at the end of 2014, but the programs' regulation and the Ministry of Education's corresponding diffusion campaign only came out in July 2015, four months before the November 2015 ECE. In our survey to teachers, of those who taught 8^{th} grade and knew about BE in 2015, 39% reported that they heard about the program in the first trimester, 26% in the second, and 33% in the third (and 2% could not remember when they found out). The programs implemented by Muralidharan and Sundararaman (2011), Glewwe et al. (2010) and Lavy (2009) were announced 7-8 months before students were tested. Even though these papers find positive and sizable effects in this short time frame, teachers in Peru might have had less time to react, especially those who found out about the program later in the year. Furthermore, when asked whether they thought there was enough time for students to improve their performance in the ECE in 2015 from the moment they found out about BE until the test, 67% of 8^{th} grade teachers who knew about the BE answered that there was not enough time.[58]

[57] One of the questions in our survey shows a hypothetical ranking of 20 urban schools from the same school district, and asks teachers to identify what position would be held by a school with the same characteristics as the one they work for, and how that position would change if every teacher in their school dedicated an extra hour a day to improve the performance of their students (extra tutoring sessions, training sessions, etc.). In 47 % of cases, math and language teachers in 8^{th} grade answered that their school would still be below the 80^{th} percentile (i.e., would not win the bonus) after everyone changed their pedagogical practices.

[58] While this could be an ex-post justification, we cannot discard that lack of time is one of the reasons the program had no impact. The future analysis of students' performance in the 2016 wave of the BE, for which teachers have the entire school year to prepare, might allow us to elucidate if this is potentially one of the reasons why the program had no effect on 8^{th} graders' performance in 2015.

8 Conclusion

Can tying teachers' pay to the performance of their students improve their learning? We examine the impact of a collective teacher pay-for-performance program (*Bono Escuela*) implemented in 2015 in all public secondary schools in Peru, and find that it had no impact on students' math and language internal grades. Our coefficients are precisely estimated, allowing us to reject effects larger than 0.017 standard deviations, well below those previously found in the literature. Moreover, we find no evidence that the teacher incentive program had differential effects over schools or students of certain characteristics. We stipulate that the lack of increase in student learning might have been triggered by certain aspects of the evaluation linked to the bonus (students' low stakes and teachers' inexperience with it). These factors, along with schools' uncertainty about their potential ranking might have discouraged teachers from exerting higher effort. Finally, we argue that the program's timing might have played a role, possibly leaving teachers with insufficient time to instill significant learning gains in their students.

All in all, the results from our study suggest that successfully scaling up teacher pay-for-performance requires a deeper understanding about the role played by the different characteristics of these programs in their success. In particular, our findings raise the question of whether the interaction between between teachers' incentives and their information is important for these programs to work. If these complementarities exist, the efficacy of teacher incentives might depend on whether they are paired with teacher training. This paper also points to the fact that the type of exam being incentivized, and particularly the stakes that students have in it, might be important for teacher pay-for-performance programs to raise student learning. Going forward, research on teacher incentives should experimentally examine the complementarities between teachers' incentives, their knowledge, and their students' stakes in the incentivized outcome.

The fact that BE had no effect in the short-term does not imply that the program would have the same learning impacts if extended for a longer period. For one, teachers would acquire more experience with both the ECE and the BE. Consequently, some of the potential issues that could be diluting the effect of the incentive program may disappear. For instance, teachers would have more time to react to the incentive, would be more familiar with the test, and schools would have more information about their potential ranking within the group of schools they are competing against. Furthermore, teachers might only find it worthwhile to make sizable investments in improving their pedagogy if the program is continued and not only a one-off event. On the other hand, the program could have undesirable long-run impacts if teachers become more acquainted with how to teach-to-the-test, or if schools divert resources away from students not reached by the ECE. Extending the program could also result in schools devoting higher effort to improving the learning of 7^{th} grade students, in anticipation of their participation in the ECE standardized test in the following year. We plan to study these issues in our future research, once students' achievement data from 2016 becomes available. Finally, the program could affect the quality of teachers attracted to public schools, impacting the performance of students in the entire school. Although there is some evidence on the role of financial incentives in shaping the attributes of candidates for public sector jobs (e.g., Dal Bó et al., 2013 and Deserranno, 2016),[59] this question has not been tackled in the context of teacher pay-for-performance programs yet.

[59] While Dal Bó et al. (2013) find that higher wages for advertised government jobs in Mexico attract candidates with higher capabilities and greater motivation for working in the public sector, Deserranno (2016) finds that higher financial incentives for health promoters in Uganda attract more candidates, but hamper retention and performance because people drawn to the position are less likely to have pro-social preferences.

References

Azmat, Ghazala and Nagore Iriberri, "The Importance of Relative Performance Feedback Information: Evidence From a Natural Experiment using High School Students," *Journal of Public Economics*, 2010, *94* (7), 435–452.

Baird, Jo-Anne, "What's in a Name? Experiments with Blind Marking in A-Level Examinations," *Educational Research*, 1998, *40* (2), 191–202.

Baker, George, "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources*, 2002, pp. 728–751.

Baker, George P, "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 1992, pp. 598–614.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Preferences and the Response To Incentives: Evidence From Personnel Data," *Quarterly Journal of Economics*, 2005, pp. 917–962.

_ , _ , and _ , "Team Incentives: Evidence From a Firm Level Experiment," *Journal of the European Economic Association*, 2013, *11* (5), 1079–1114.

_ , _ , _ et al., "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment," *Quarterly Journal of Economics*, 2007, *122* (2), 729–773.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton, "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India," 2016. NBER Working Paper 22746.

Behrman, Jere R, Susan W Parker, Petra E Todd, and Kenneth I Wolpin, "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools," *Journal of Political Economy*, 2015, *123* (2), 325–364.

Bó, Ernesto Dal, Frederico Finan, and Martín A Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.

Botelho, Fernando, Ricardo A Madeira, and Marcos A Rangel, "Racial Discrimination in Grading: Evidence from Brazil," *American Economic Journal: Applied Economics*, 2015, *7* (4), 37–52.

Bruns, Barbara and Javier Luque, *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*, World Bank Publications, 2015.

_ , Deon Filmer, and Harry Anthony Patrinos, *Making Schools Work: New Evidence on Accountability Reforms*, World Bank Publications, 2011.

Burgess, Simon and Ellen Greaves, "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities," *Journal of Labor Economics*, 2013, *31* (3), 535–576.

Calsamiglia, Caterina and Annalisa Loviglio, "Maturity and School Outcomes in an Inflexible System: Evidence from Catalonia," 2016. Working Paper.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers, "Missing in Action :Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 2006, *20* (1), 91–116.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104* (9), 2633–2679.

Chevalier, Judith and Glenn Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 1997, *105* (6), 1167–1200.

Chong, Alberto, Isabelle Cohen, Erica Field, Eduardo Nakasone, and Maximo Torero, "Iron Deficiency and Schooling Attainment in Peru," *American Economic Journal: Applied Economics*, 2016.

Contreras, Dante and Tomás Rau, "Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile," *Economic Development and Cultural Change*, 2012, *61* (1), 219–246.

Deaton, Angus and Nancy Cartwright, "Understanding and Misunderstanding Randomized Controlled Trials," 2016. National Bureau of Economic Research Working Paper w22595.

Deserranno, Erika, "Financia Incentives as Signals: Experimental Evidence From the Recruitment of Health Workers," 2016. Working Paper.

Ewijk, Reyn Van, "Same Work, Lower Grade? Student Ethnicity and Teachers Subjective Assessments," *Economics of Education Review*, 2011, *30* (5), 1045–1058.

Figlio, David N, "Testing, crime and punishment," *Journal of Public Economics*, 2006, *90* (4), 837–851.

_ **and Joshua Winicki**, "Food for thought: the effects of school accountability plans on school nutrition," *Journal of Public Economics*, 2005, *89* (2), 381–394.

Fryer, Roland G, "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," *Journal of Labor Economics*, 2013, *31* (2), 373–407.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2010, *2* (3), 205–227.

Gneezy, Uri and Aldo Rustichini, "A Fine is a Price," *Journal of Legal Studies*, 2000, *29* (1).

_ **and** _ , "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 2000, pp. 791–810.

_ **, Stephan Meier, and Pedro Rey-Biel**, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 2011, *25* (4), 191–209.

Goodman, Sarena F and Lesley J Turner, "The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program," *Journal of Labor Economics*, 2013, *31* (2), 409–420.

Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton, "Autonomy and Incentives in Chinese State Enterprises," *Quarterly Journal of Economics*, 1994, pp. 183–209.

Hanna, Rema N and Leigh L Linden, "Discrimination in Grading," *American Economic Journal: Economic Policy*, 2012, *4* (4), 146–168.

Hanushek, Eric A and Steven G Rivkin, "Generalizations about Using Value-Added Measures of Teacher Quality," *American Economic Review*, 2010, *100* (2), 267–271.

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 1982, pp. 324–340.

_ **and Paul Milgrom**, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 1991, *7*, 24–52.

Imberman, Scott A and Michael F Lovenheim, "Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teacher Incentive Pay System," *Review of Economics and Statistics*, 2015, *97* (2), 364–386.

Jacob, Brian A and Steven D Levitt, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 2003, pp. 843–877.

Kandel, Eugene and Edward P Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy*, 1992, pp. 801–817.

Kandori, Michihiro, "Social Norms and Community Enforcement," *Review of Economic Studies*, 1992, *59* (1), 63–80.

Kerwin, Jason T and Rebecca Thornton, "Making the Grade: Understanding What Works for Teaching Literacy in Rural Uganda," 2015.

Koretz, Daniel M, "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity," *Journal of Human Resources*, 2002, *37* (4), 752–777.

Lavy, Victor, "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, 2002, *110* (6), 1286–1317.

_ , "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence From a Natural Experiment," *Journal of Public Economics*, 2008, *92* (10), 2083–2105.

_ , "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, 2009, *99* (5), 1979–2021.

_ , "Teachers' Pay for Performance in the Long-Run: Effects on Students' Educational and Labor Market Outcomes in Adulthood," 2015. NBER Working Paper 20983.

Lazear, Edward P and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Mizala, Alejandra and Ben Ross Schneider, "Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010)," *Governance*, 2014, *27* (1), 87–109.

_ **and Hugo Ñopo**, "Measuring the Relative Pay of School Teachers in Latin America 1997–2007," *International Journal of Educational Development*, 2016, *47*, 20–32.

Muralidharan, Karthik and Venkatesh Sundararaman, "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India," *Economic Journal*, 2010, *120* (546), F187–F203.

_ **and** _ , "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Neal, Derek, "The Design of Performance Pay in Education," *Handbook of the Economics of Education*, 2011, *4*, 495–550.

_ **and Diane Whitmore Schanzenbach**, "Left Behind by Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Newstead, Stephen E and Ian Dennis, "Blind Marking and Sex Bias in Student Assessment," *Assessment and Evaluation in higher Education*, 1990, *15* (2), 132–139.

Oyer, Paul, "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics*, 1998, pp. 149–185.

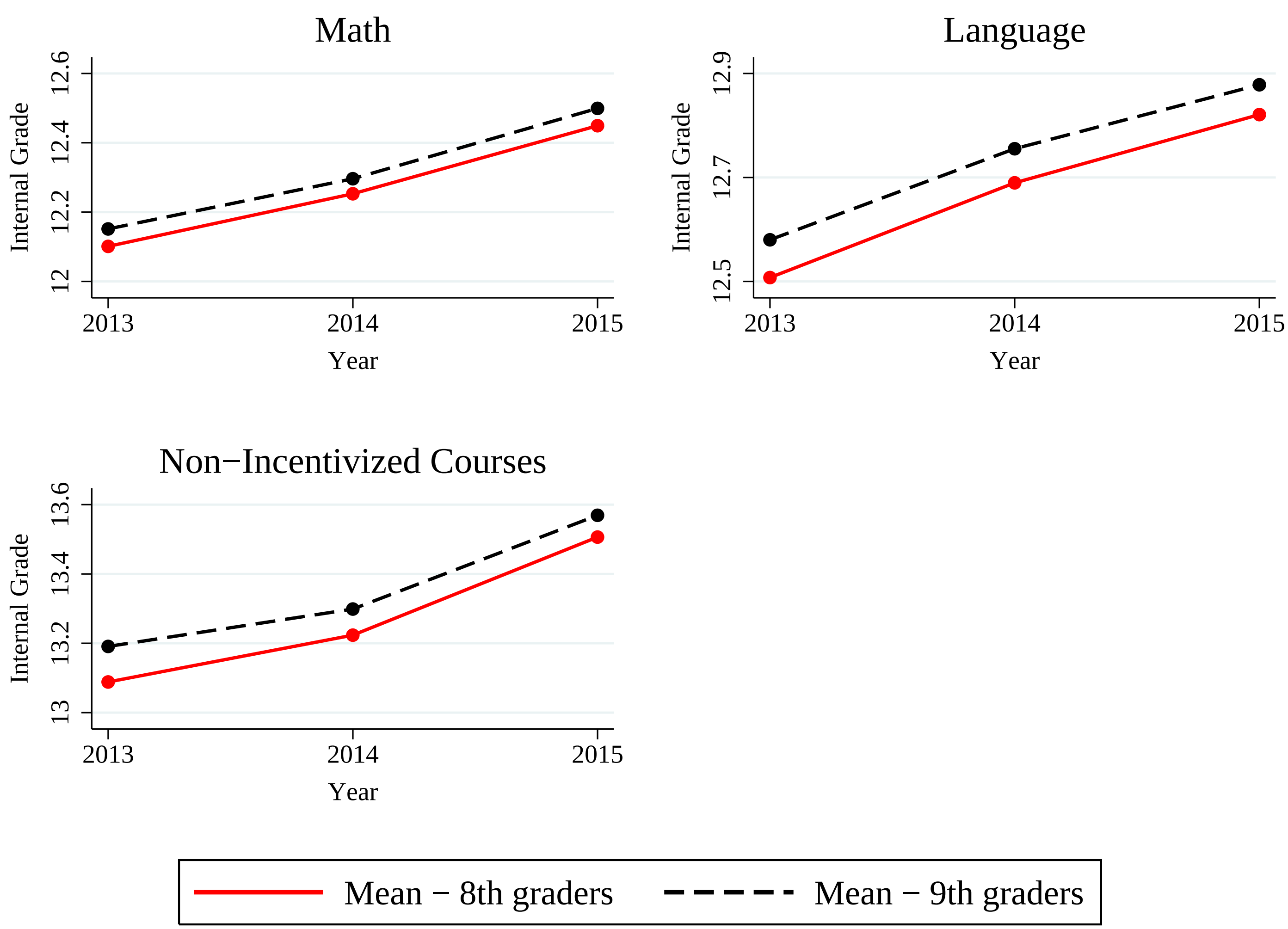
Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, 1999, *37* (1), 7–63.

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

Springer, Matthew G, Laura Hamilton, Daniel F McCaffrey, Dale Ballou, Vi-Nhuan Le, Matthew Pepper, JR Lockwood, and Brian M Stecher, "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching.," *National Center on Performance Incentives*, 2010.

Tran, Anh and Richard Zeckhauser, "Rank as an Inherent Incentive: Evidence From a Field Experiment," *Journal of Public Economics*, 2012, *96* (9), 645–650.

Figure 1: Trend in Average Internal Grades for 8^{th} and 9^{th} Graders



Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools which were eligible for taking the 2015 ECE standardized test and which are registered in the Ministry of Education's SIAGIE administrative system. The figures plot the average of all 8^{th} and 9^{th} graders internal grades in math, language and non-incentivized courses, respectively. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce.

Figure 2: Timing of BE in Secondary Schools

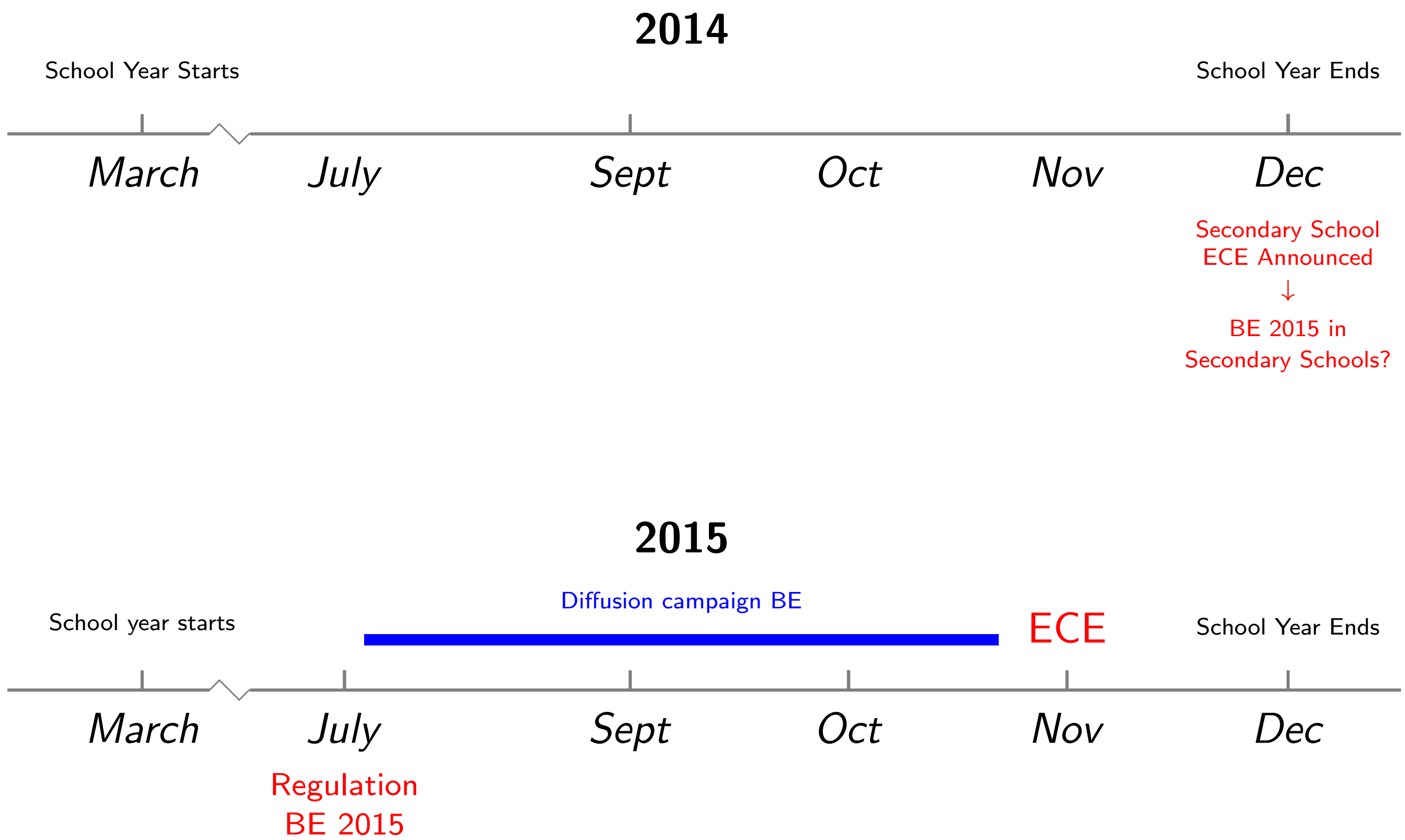
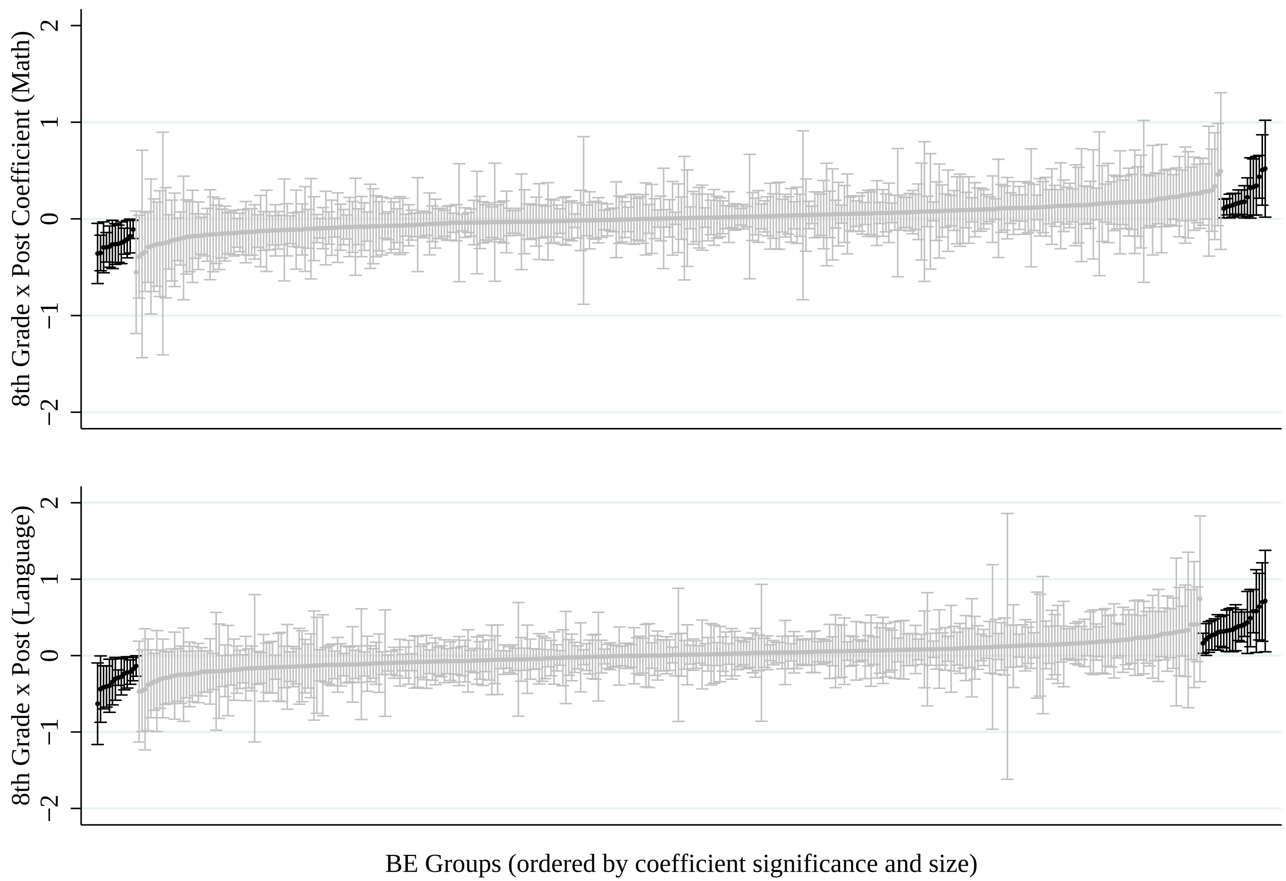


Figure 3: Effect of Teacher Incentive on Students' Math and Language Internal Grades in each BE Group



Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools which were eligible for taking the 2015 ECE standardized test and which are registered in the Ministry of Education's SIAGIE administrative system. The figures plot the 8^{th} *Grade* x *Post* coefficients and their 95% confidence intervals separately estimated for each BE group in math and language, respectively. BE groups in both figures are ordered by significance and coefficient size, and the ordering is separately done in each figure.

Table 1: Assignment of Score in BE

Weight	Indicator	Relevant Grades
40%	Average math and language score in 2015 ECE standardized tests	8th Grade
35%	Intra-annual retention rates	All Grades
5%	Enrollment of students in SIAGIE administrative system	All Grades
12%	Teacher attendance, management of school infrastructure and compliance with class hours	All Grades
8%	Pedagogical practices and learning environment	8th Grade

Source: Decree 203-2015.

Table 2: Summary Statistics for 8^{th} and 9^{th} Graders

	8^{th} Grade		9^{th} Grade	
	Mean	Std. Dev	Mean	Std. Dev
Final Grade (0-20)				
Math	12.27	2.17	12.32	2.17
Language	12.67	2.07	12.74	2.07
Other courses - average	13.27	1.60	13.35	1.60
Passed the Course				
Math	0.84	0.37	0.85	0.36
Language	0.89	0.31	0.90	0.30
Other courses - average	0.93	0.15	0.94	0.14
Other Individual Characteristics				
Male	0.49	0.50	0.50	0.50
Repeated last year	0.06	0.23	0.04	0.20
Foreigner	0.00	0.05	0.00	0.05
Spanish is native tongue	0.84	0.37	0.83	0.38
Has a disability	0.00	0.06	0.00	0.06
Father is alive	0.90	0.30	0.89	0.31
Mother is alive	0.97	0.16	0.97	0.17
Father lives in HH	0.76	0.43	0.77	0.42
Mother lives in HH	0.80	0.40	0.80	0.40
Number of students	1,090,496		1,018,310	
Grade/School Characteristics				
Rural	0.41	0.49	0.40	0.49
Number of classes	2.00	1.92	1.94	1.84
Teacher-pupil ratio	19.61	8.72	18.98	8.75
Number of teachers	10.74	6.55	10.95	6.71
% of teachers instructing the other grade	0.83	0.22	0.83	0.20
Average age of teachers	41.64	5.34	41.66	5.26
% of male teachers	0.60	0.21	0.60	0.20
Number of school-year observations	23,810		23,469	

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. We exclude students for which we have no grades and/or no individual controls (0.4%). Since teacher data is missing for a small subsample of schools, the number of grade-observations for teacher characteristics is 23,462 and 23,127 in 8^{th} and 9^{th} grade. *Final Grade* is the students' internal grades at the end of the school year in math, language and non-incentivized courses. *Passed the Course* is a dummy for whether the student got an 11 or higher in that particular course (the requirement for passing). We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. *Repeated last year* is a dummy for whether the student was retained in the same grade at the end of the previous year. *Rural* is a dummy for whether the school is in a rural area, and *Number of classes* is the number of classes in the student's grade and year. *Number of teachers* is the total number of teachers in that grade and year, and *% of teachers instructing the other grade* is the % of 8^{th} (9^{th}) teachers also teaching 9^{th} (8^{th}) grade in the same school.

Table 3: Effect of Teacher Incentive on Students' Math and Language Internal Grades

	Math			Language		
	(1)	(2)	(3)	(4)	(5)	(6)
8^{th} Grade x Post	-0.001	-0.003	-0.005	0.006	0.004	0.001
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
8^{th} Grade	-0.022***	-0.016***	-0.008**	-0.034***	-0.026***	-0.016***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Repeated last year			-0.570***			-0.623***
			(0.006)			(0.007)
Male			0.115***			0.334***
			(0.003)			(0.003)
Foreigner			0.046**			0.089***
			(0.019)			(0.019)
Spanish is native tongue			0.171***			0.174***
			(0.007)			(0.007)
Has a disability			-0.262***			-0.256***
			(0.014)			(0.015)
Father is alive			0.066***			0.066***
			(0.004)			(0.003)
Mother is alive			0.040***			0.049***
			(0.006)			(0.006)
Father lives in HH			0.026***			0.020***
			(0.002)			(0.002)
Mother lives in HH			0.014***			0.019***
			(0.003)			(0.002)
Observations	2108806	2108806	2108806	2108793	2108793	2108793
R^2	0.000	0.071	0.092	0.000	0.087	0.135
Year FE	X	X	X	X	X	X
School FE		X	X		X	X
Individual Controls			X			X

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Effect of Teacher Incentive on Students' Internal Grades in Non-Incentivized Courses

	Non-Incentivized Courses		
	(1)	(2)	(3)
8^{th} Grade x Post	0.014***	0.014***	0.011***
	(0.004)	(0.004)	(0.004)
8^{th} Grade	-0.043***	-0.038***	-0.028***
	(0.003)	(0.003)	(0.002)
Repeated last year			-0.601***
			(0.006)
Male			0.284***
			(0.003)
Foreigner			0.047***
			(0.015)
Spanish is native tongue			0.120***
			(0.005)
Has a disability			-0.232***
			(0.012)
Father is alive			0.063***
			(0.003)
Mother is alive			0.044***
			(0.005)
Father lives in HH			0.022***
			(0.002)
Mother lives in HH			0.013***
			(0.002)
Observations	2108972	2108972	2108972
R^2	0.001	0.120	0.185
Year FE	X	X	X
School FE		X	X
Individual Controls			X

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable is students' internal grades in non-incentivized courses, standardized by course-year. We first standardize each of the non-incentivized courses by course-year, and then take the average. Non-incentivized courses are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Heterogeneous Effect of Teacher Incentive on Students' Internal Grades by Schools' Socioeconomic Status (SES) Rank

	Within-BE Group Percentile by SES Index					
	$55^{th}-100^{th}$	$65^{th}-100^{th}$	$75^{th}-100^{th}$	$55^{th}-95^{th}$	$65^{th}-95^{th}$	$75^{th}-95^{th}$
Panel A: Math						
8th Grade x Post	-0.019**	-0.012	-0.006	-0.019**	-0.012	-0.006
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
8th Grade x Post x I(Percentile)	0.024*	0.014	0.004	0.028**	0.019	0.009
	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.016)
8th Grade x Post x Top 5%				0.003	-0.004	-0.009
				(0.024)	(0.024)	(0.024)
Observations	2108570	2108570	2108570	2108570	2108570	2108570
R^2	0.092	0.092	0.092	0.092	0.092	0.092
P-Value (Percentiles)	0.806	0.724	0.637	0.805	0.708	0.640
P-Value (Top 5%)				0.872	0.855	0.751
Panel B: Language						
8th Grade x Post	-0.011	-0.008	-0.004	-0.010	-0.008	-0.004
	(0.011)	(0.010)	(0.009)	(0.011)	(0.010)	(0.009)
8th Grade x Post x I(Percentile)	0.020	0.018	0.015	0.028*	0.030*	0.031
	(0.016)	(0.016)	(0.018)	(0.016)	(0.017)	(0.020)
8th Grade x Post x Top 5%				-0.023	-0.026	-0.029
				(0.031)	(0.031)	(0.030)
Observations	2108557	2108557	2108557	2108557	2108557	2108557
R^2	0.135	0.135	0.135	0.135	0.135	0.135
P-Value (Percentiles)	0.777	0.995	0.717	0.419	0.493	0.569
P-Value (Top 5%)				0.065	0.038	0.020

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables in Panels A and B are students' internal grades in math and language, respectively, standardized by course-year. We construct a SES index (taking values 0-3) for each 8^{th} student in 2015, adding up three dummies for whether his first language is Spanish, and whether his mother and father have more than a primary school education. We then calculate the average index for each school, and stratify schools according to their ranking within their BE group. *I(Percentile)* is a dummy for whether the school belongs to the percentile indicated in the column header in his BE group, and *Top 5%* is a dummy for whether the school is above the 95^{th} percentile in his BE group. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. All regressions include year and school fixed effects, and the standard individual controls. Regressions in columns (1)-(3) include the three-way interaction between 8^{th} *Grade*, *Post* and *I(Percentile)*, and those in columns (4)-(6) also include the three-way interaction between 8^{th} *Grade*, *Post* and *Top 5%*. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Heterogeneous Effect of Teacher Incentive on Students' Internal Grades

	Male	Spanish Speaker	Parents High Educ	Repeated	Lagged Grade	Ln Enrollment	Num. Classes	Rural
Panel A: Math Grades								
8^{th} Grade x Post	-0.010	-0.004	-0.005	-0.003	-0.002	0.033	-0.002	-0.007
	(0.008)	(0.012)	(0.007)	(0.007)	(0.008)	(0.023)	(0.009)	(0.008)
8^{th} Grade x Post x Covariate	0.010	-0.000	-0.003	-0.019	0.006	-0.008	-0.001	0.016
	(0.009)	(0.013)	(0.005)	(0.016)	(0.005)	(0.006)	(0.002)	(0.012)
Observations	2108806	2108806	1851727	2108806	1382813	2108806	2108806	2108806
R^2	0.092	0.092	0.099	0.092	0.440	0.092	0.092	0.092
Panel B: Language Grades								
8^{th} Grade x Post	0.006	-0.000	-0.004	0.002	0.005	-0.023	-0.008	0.004
	(0.009)	(0.013)	(0.009)	(0.008)	(0.009)	(0.027)	(0.010)	(0.009)
8^{th} Grade x Post x Covariate	-0.010	0.002	0.003	-0.008	0.001	0.006	0.002	-0.017
	(0.010)	(0.015)	(0.006)	(0.016)	(0.005)	(0.007)	(0.002)	(0.014)
Observations	2108793	2108793	1851715	2108793	1382756	2108793	2108793	2108793
R^2	0.135	0.135	0.142	0.135	0.417	0.135	0.135	0.135

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. Heterogeneities by *Lagged Grade* exclude the year 2013 for which students' previous grade is unavailable, and heterogeneities by parents' education exclude 12% of students in 2013-2015 for which this variable is missing. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, and *Covariate* is the variable indicated in the column header. All regressions include school and year fixed effects, as well as the standard individual controls and the three-way interaction between 8^{th} *Grade*, *Post* and *Covariate*. We only report two coefficients for exposition purposes. *Spanish Speaker* is a dummy for whether the student's first language is Spanish, and *Parents High Educ* is 0 if both parents have a primary school degree or less, is 1 if only one of the parents has more than a primary school degree, and 2 if both. *Repeated* is a dummy for whether the student was retained in the same grade at the end of the previous year, and *Lagged Grade* is the students' internal grade in that particular course in the previous year, standardized by school and grade. *Ln Enrollment* is the log of the number of students enrolled in that year and grade. *Num. Classes* is the number of classes in the student's grade and year, *Rural* is a dummy for whether the school is in a rural area. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Test for Parallel Trends in Students' Internal Grades

	Math	Language	Non-Incentivized Courses
8^{th} Grade x Post	-0.004	0.002	0.016***
	(0.007)	(0.009)	(0.005)
8^{th} Grade x 2014	0.002	0.001	0.010**
	(0.007)	(0.009)	(0.004)
8^{th} Grade	-0.009*	-0.017***	-0.033***
	(0.005)	(0.007)	(0.003)
Observations	2108806	2108793	2108972
R^2	0.092	0.135	0.185

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE are registered in the Ministry of Education's SIAGIE administrative system. All regressions include year fixed effects, school fixed effects, and the standard controls. The dependent variables are students' internal grades in math, language and non-incentivized courses, standardized by course-year. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, and *2014* is a dummy for the year 2014. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Heterogeneity by Overlap of 8^{th} and 9^{th} Grade Teachers

	Math			Language		
	Low	Med	High	Low	Med	High
8^{th} Grade x Post	-0.014	-0.014	0.013*	0.010	0.001	-0.009
	(0.014)	(0.012)	(0.007)	(0.017)	(0.014)	(0.008)
8^{th} Grade	-0.001	-0.006	-0.016***	-0.017	-0.017**	-0.014***
	(0.008)	(0.007)	(0.004)	(0.011)	(0.008)	(0.005)
Repeated last year	-0.594***	-0.533***	-0.582***	-0.637***	-0.587***	-0.642***
	(0.013)	(0.008)	(0.009)	(0.014)	(0.009)	(0.010)
Male	0.123***	0.128***	0.097***	0.369***	0.367***	0.275***
	(0.006)	(0.005)	(0.004)	(0.007)	(0.005)	(0.005)
Foreigner	0.048*	0.017	0.124**	0.084***	0.082***	0.142**
	(0.027)	(0.028)	(0.055)	(0.027)	(0.029)	(0.057)
Spanish is native tongue	0.130***	0.180***	0.198***	0.129***	0.185***	0.202***
	(0.013)	(0.012)	(0.009)	(0.013)	(0.013)	(0.009)
Has a disability	-0.231***	-0.247***	-0.300***	-0.200***	-0.243***	-0.310***
	(0.028)	(0.025)	(0.019)	(0.026)	(0.027)	(0.021)
Father is alive	0.070***	0.068***	0.061***	0.073***	0.061***	0.065***
	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)
Mother is alive	0.020	0.044***	0.055***	0.029**	0.051***	0.067***
	(0.013)	(0.010)	(0.008)	(0.011)	(0.010)	(0.009)
Father lives in HH	0.034***	0.032***	0.011***	0.028***	0.023***	0.006
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Mother lives in HH	0.013***	0.012***	0.020***	0.015***	0.017***	0.025***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	740453	672268	696085	740446	672275	696072
R^2	0.072	0.084	0.126	0.116	0.129	0.166
Avg. % of teachers in both grades	0.392	0.654	0.906	0.392	0.654	0.906

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. Columns Low, Med and High restrict the sample to students in schools with a low, medium and high average overlap between 8^{th} and 9^{th} grade teachers in 2013-2015. Overlap between 8^{th} and 9^{th} grade teachers is the % of teachers in 8^{th} grade also instructing in 9^{th} grade. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: Test for Changes in Teacher Composition Across Grades

	Average Age	% Male	Average Number of Classes	Average Number of Schools	% New to School	% New to School-Grade
8^{th} Grade x Post	-0.013	0.001	0.016	-0.005**	-0.000	-0.000
	(0.021)	(0.001)	(0.016)	(0.002)	(0.001)	(0.002)
8^{th} Grade	0.043***	-0.006***	-0.090***	0.000	0.001	0.001
	(0.013)	(0.001)	(0.011)	(0.001)	(0.001)	(0.001)
Observations	46614	46615	46615	46615	31332	31332
R^2	0.869	0.738	0.953	0.737	0.808	0.758
Mean Dep. Variable	41.685	0.598	11.790	1.634	0.474	0.552

Notes: The sample includes all public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE standardized test, registered in the Ministry of Education's SIAGIE administrative system, and with data on teacher characteristics. The unit of analysis in these regressions is a school-grade-year, for 8^{th} and 9^{th} grade. *Average Age* is the average age of teachers in that grade, and *% Male* is the % of teachers in that grade that are male. *Average Number of Classes* is the average number of courses taught by teachers in that grade, and *Average Number of Schools* is the mean number of different secondary schools in which the teacher works. *% New to School-Grade* are the proportion of teachers in that particular grade who are new to the school, or new to that particular grade , respectively. All regressions include school fixed effects, year fixed effects, a dummy for 8^{th} *Grade*, and the interaction between 8^{th} *Grade* and a dummy for 2015 (i.e., *Post*). The regressions in columns 5 and 6 do not include the year 2013 since we do not have information on teachers' appointments in 2012. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 10: Effect of Teacher Incentive by Average Salary, Number of Classes, School's Experience with Primary School BE and BE Group Size

	Math					Language				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
8^{th} Grade x Post	0.249	-0.011	-0.006	-0.007	-0.014*	0.299	0.000	0.005	-0.000	-0.010
	(0.501)	(0.008)	(0.009)	(0.007)	(0.008)	(0.585)	(0.010)	(0.011)	(0.009)	(0.011)
8^{th} Grade x Post x Ln (Average Salary)	-0.035					-0.041				
	(0.069)					(0.081)				
8^{th} Grade x Post x One Class		0.030***					0.004			
		(0.011)					(0.013)			
8^{th} Grade x Post x BE Primary			0.003					-0.007		
			(0.013)					(0.016)		
8^{th} Grade x Post x BE Primary Winner				0.016					0.010	
				(0.018)					(0.024)	
8^{th} Grade x Post x Small BE Group					0.026**					0.033**
					(0.013)					(0.016)
Observations	2077227	2108806	2108806	2108806	2108806	2077214	2108793	2108793	2108793	2108793
R^2	0.092	0.092	0.092	0.092	0.092	0.135	0.135	0.135	0.135	0.135
P-Value (sum of both coefficients)	0.621	0.008	0.739	0.568	0.237	0.601	0.558	0.893	0.669	0.062

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, *Ln (Average Salary)* is the average salary of teachers in each school in 2015 (in logs), obtained from the 2015 school census, and *One Class* is a dummy for whether the student attends a school in which there is only one class in his grade. *BE Primary* is a dummy for whether a primary school that participated in the BE in the past operates in the same building, and *BE Primary* is a dummy for whether there is a primary school in the building that won the BE bonus in the past. *Small BE Group* is a dummy for whether the number of schools in the corresponding BE group is below the median. All regressions include year and school fixed effects, the standard individual controls, and the three-way interaction between 8^{th} *Grade*, *Post* and the specific heterogeneity variable. We only report two coefficients for exposition purposes. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 11: Effect of Teacher Incentive on Teachers' Pedagogical Practices

	8^{th} Grade	Other Grades	Difference	P-Value
Panel A: All Math/Language Teachers				
Improved attendance	0.207	0.157	0.050**	0.024
More homework, evaluations and/or tutoring sessions	0.471	0.370	0.101***	0.000
Paid more attention to weakest students	0.683	0.637	0.046*	0.097
Training programs or feedback sessions	0.548	0.542	0.006	0.828
Increased difficulty of classes	0.192	0.135	0.056***	0.007
Decreased difficulty of classes	0.148	0.138	0.010	0.620
More multiple choice tests	0.385	0.299	0.086***	0.002
Other	0.130	0.150	-0.020	0.317
Number of teachers	454	865		
Panel B: Math/Language Teachers in Both Grades				
Improved attendance	0.203	0.143	0.060***	0.000
More homework, evaluations and/or tutoring sessions	0.460	0.326	0.134***	0.000
Paid more attention to weakest students	0.677	0.657	0.020	0.209
Training programs or feedback sessions	0.523	0.494	0.029	0.149
Increased difficulty of classes	0.197	0.157	0.040**	0.016
Decreased difficulty of classes	0.146	0.131	0.014	0.298
More multiple choice tests	0.391	0.337	0.054**	0.017
Other	0.123	0.143	-0.020	0.250
Number of teachers	350	350		

Notes: The sample includes all survey respondents who taught math or language in 8^{th} and other grades in 2015, and knew about the BE program during the 2015 academic year. Panel B only includes those who taught math or language in 8^{th} grade and other grades. Teachers were asked whether they changed their pedagogical practices in 2015 as a result of BE, and could answer any of the options specified in the table rows. We asked them separately about changes while teaching 8^{th} grade (column 1) as opposed to all other grades (column 2), in case the teacher taught both. * significant at 10%; ** significant at 5%; *** significant at 1%

9 Appendix Figures and Tables

Table A.1: Test for Parallel Trends Comparing Public and Private Schools

	Math	Language	Non-Incentivized Courses
Public x Post	0.116***	0.100***	0.143***
	(0.010)	(0.011)	(0.007)
Public x 2014	0.044***	0.063***	0.036***
	(0.008)	(0.010)	(0.006)
Repeated last year	-0.527***	-0.586***	-0.548***
	(0.006)	(0.006)	(0.005)
Male	0.109***	0.323***	0.271***
	(0.003)	(0.003)	(0.002)
Foreigner	0.037***	0.058***	0.061***
	(0.012)	(0.011)	(0.009)
Spanish is native tongue	0.183***	0.193***	0.126***
	(0.007)	(0.007)	(0.005)
Has a disability	-0.247***	-0.265***	-0.232***
	(0.014)	(0.015)	(0.013)
Father is alive	0.073***	0.075***	0.068***
	(0.003)	(0.004)	(0.003)
Mother is alive	0.031***	0.038***	0.039***
	(0.006)	(0.006)	(0.005)
Father lives in HH	0.037***	0.031***	0.030***
	(0.002)	(0.002)	(0.002)
Mother lives in HH	0.011***	0.016***	0.011***
	(0.002)	(0.002)	(0.002)
Observations	1514619	1514593	1514717
R^2	0.155	0.196	0.293

Notes: The sample includes all 8^{th} grade students in 2013-2015, in public and private schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. All regressions include year fixed effects, school fixed effects, and the standard controls. The dependent variables are students' internal grades in math, language and non-incentivized courses, standardized by course-year. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. *Public* is a dummy for whether the students attends a public school, *Post* is a dummy for the year 2015, and *2014* is a dummy for the year 2014. The coefficient for the *Public* dummy is not display since this variable is perfectly collinear with the corresponding school fixed effect. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.2: Effect of Teacher Incentive on Students' Grades in Non-Incentivized Courses

	Arts	Science	Social Studies	English	Civics	Human Relations	Physical Education	Religion	Educ. for the Workforce
8^{th} Grade x Post	0.003	-0.001	0.017*	0.011	0.008	0.019**	0.012	0.014*	0.012
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
8^{th} Grade	-0.037***	0.052***	-0.039***	-0.046***	-0.055***	-0.044***	-0.016***	-0.042***	-0.027***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Repeated last year	-0.621***	-0.570***	-0.602***	-0.573***	-0.598***	-0.588***	-0.617***	-0.611***	-0.624***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.008)
Male	0.336***	0.236***	0.259***	0.281***	0.339***	0.384***	0.074***	0.393***	0.256***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Foreigner	0.029*	0.057***	0.039**	0.163***	0.041**	0.030	0.061***	-0.015	0.003
	(0.017)	(0.018)	(0.020)	(0.019)	(0.020)	(0.018)	(0.015)	(0.018)	(0.018)
Spanish is native tongue	0.088***	0.148***	0.128***	0.147***	0.128***	0.138***	0.092***	0.103***	0.107***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
Has a disability	-0.188***	-0.261***	-0.229***	-0.276***	-0.235***	-0.239***	-0.257***	-0.185***	-0.219***
	(0.016)	(0.014)	(0.016)	(0.014)	(0.015)	(0.014)	(0.015)	(0.016)	(0.016)
Father is alive	0.064***	0.063***	0.065***	0.068***	0.063***	0.066***	0.052***	0.062***	0.064***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Mother is alive	0.050***	0.043***	0.040***	0.050***	0.038***	0.043***	0.047***	0.039***	0.043***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Father lives in HH	0.024***	0.026***	0.024***	0.020***	0.025***	0.023***	0.013***	0.022***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother lives in HH	0.012***	0.014***	0.013***	0.015***	0.014***	0.014***	0.012***	0.016***	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	2108795	2108792	2108780	2108791	2108793	2108777	2108794	2071576	2108778
R^2	0.196	0.127	0.138	0.143	0.154	0.168	0.235	0.183	0.186

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.3: Non-Linear Heterogeneous Effects by Students' Lagged Grade

	Math	Language
Panel A: Lagged Grade Quartiles		
8^{th} Grade x Post	-0.003	0.003
	(0.009)	(0.011)
8^{th} Grade x Post x Q2	-0.006	0.009
	(0.009)	(0.010)
8^{th} Grade x Post x Q3	-0.010	0.008
	(0.010)	(0.011)
8^{th} Grade x Post x Q4	0.014	0.001
	(0.014)	(0.015)
Observations	1382813	1382756
R^2	0.393	0.387
P-value (sum of coefficients Q2)	0.346	0.281
P-value (sum of coefficients Q3)	0.198	0.349
P-value (sum of coefficients Q4)	0.380	0.792
Panel B: Lagged Grade Terciles		
8^{th} Grade x Post	-0.004	-0.004
	(0.009)	(0.010)
8^{th} Grade x Post x T2	-0.014	0.010
	(0.009)	(0.009)
8^{th} Grade x Post x T3	0.014	0.011
	(0.013)	(0.012)
Observations	1382813	1382756
R^2	0.359	0.363
P-value (sum of coefficients T2)	0.050	0.622
P-value (sum of coefficients T3)	0.444	0.588

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2014-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. We exclude the year 2013 for which students' previous grade is unavailable. The dependent variables are students' internal grades in math and language, standardized by course-year. Students in Panel A (B) are divided into quartiles (terciles) according to their lagged grade (i.e., their internal grade in that particular course in the previous year, standardized by school and grade). 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. All regressions include school and year fixed effects, as well as the standard individual controls and the three-way interaction between 8^{th} *Grade*, *Post* and the Quartile or Tercile dummies. We only report the triple interactions for exposition purposes. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.4: Heterogeneous Effect of Teacher Incentive on Students' Internal Grades by Schools' Infrastructure Quality Rank

	Within-BE Group Percentile by Infrastructure Quality Index					
	$55^{th}-100^{th}$	$65^{th}-100^{th}$	$75^{th}-100^{th}$	$55^{th}-95^{th}$	$65^{th}-95^{th}$	$75^{th}-95^{th}$
Panel A: Math						
8th Grade x Post	0.012	0.008	0.007	0.012	0.008	0.007
	(0.011)	(0.010)	(0.009)	(0.011)	(0.010)	(0.009)
8th Grade x Post x I(Percentile)	-0.022	-0.021	-0.022	-0.017	-0.016	-0.016
	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
8th Grade x Post x Top 5%				-0.033*	-0.030	-0.029
				(0.019)	(0.019)	(0.018)
Observations	1992846	1992846	1992846	1992846	1992846	1992846
R^2	0.091	0.091	0.092	0.091	0.091	0.091
P-Value (Percentiles)	0.233	0.254	0.257	0.702	0.403	0.452
P-Value (Top 5%)				0.785	0.163	0.173
Panel B: Language						
8th Grade x Post	0.004	0.006	0.005	0.004	0.006	0.005
	(0.012)	(0.012)	(0.010)	(0.012)	(0.012)	(0.010)
8th Grade x Post x I(Percentile)	-0.004	-0.007	-0.006	-0.004	-0.008	-0.007
	(0.016)	(0.016)	(0.017)	(0.018)	(0.019)	(0.020)
8th Grade x Post x Top 5%				-0.004	-0.005	-0.004
				(0.023)	(0.023)	(0.023)
Observations	1992834	1992834	1992834	1992834	1992834	1992834
R^2	0.135	0.135	0.135	0.135	0.135	0.135
P-Value (Percentiles)	0.864	0.775	0.682	0.341	0.962	0.884
P-Value (Top 5%)				0.135	0.561	0.528

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables in Panels A and B are students' internal grades in math and language, respectively, standardized by course-year. We construct a school infrastructure quality index (taking values 0-12) from the Peruvian school census, adding the following dummy variables for whether the school building has each of these characteristics: blackboards in good conditions, brick/concrete walls, tile/concrete roof, playground, computer room, teachers' lounge, sports area, electricity, running water, sewage, internet, library, science lab. We then stratify schools according to their ranking within their BE group, and exclude BE groups in which this variable is unavailable for more than 25% of schools. *I(Percentile)* is a dummy for whether the school belongs to the percentile indicated in the column header in his BE group, and *Top 5%* is a dummy for whether the school is above the 95^{th} percentile in his BE group. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. All regressions include year and school fixed effects, and the standard individual controls. Regressions in columns (1)-(3) include the three-way interaction between 8^{th} *Grade*, *Post* and *I(Percentile)*, and those in columns (4)-(6) also include the three-way interaction between 8^{th} *Grade*, *Post* and *Top 5%*. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.5: Cross-Sectional Correlation Between Average ECE Test Scores and Internal Grades in 2015

	Math		Language	
	Average ECE (z-score)	% Very Low ECE	Average ECE (z-score)	% Very Low ECE
Average internal grade (z-score)	0.116***		0.103***	
	(0.010)		(0.008)	
Failed course (% of students)		0.071***		0.091***
		(0.018)		(0.021)
Spanish as native tongue (% of students)	0.667***	-0.180***	0.836***	-0.250***
	(0.044)	(0.012)	(0.039)	(0.012)
Mother with high education (% of students)	0.930***	-0.223***	1.143***	-0.221***
	(0.078)	(0.020)	(0.065)	(0.018)
Father with high education (% of students)	0.438***	-0.120***	0.558***	-0.175***
	(0.075)	(0.020)	(0.060)	(0.018)
Repeated last year (% of students)	-0.138	0.044	0.079	-0.053
	(0.156)	(0.038)	(0.118)	(0.039)
Male (% of students)	-0.171***	0.037***	0.086*	-0.029**
	(0.054)	(0.014)	(0.046)	(0.013)
Teacher-pupil ratio	0.010***	-0.002***	0.010***	-0.002***
	(0.001)	(0.000)	(0.001)	(0.000)
Long school-day	0.309***	-0.074***	0.187***	-0.052***
	(0.024)	(0.006)	(0.018)	(0.005)
Rural	-0.050**	0.019***	-0.135***	0.049***
	(0.023)	(0.006)	(0.018)	(0.006)
Observations	8010	8010	8010	8010
R^2	0.501	0.491	0.684	0.617

Notes: The sample includes all public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable in columns 1 and 3 are 8^{th} graders' average ECE grades in math and language, expressed as a z-score. The dependent variable in columns 2 and 4 is the % of students with very low achievement in the 2015 ECE. *Average internal grade (z-score)* is the school's average internal grade for 8^{th} grade students in 2015 , standardized across schools, in math (column 1) and language (column 3). *Failed course* measures the % of 8^{th} graders that failed math (column 2) and language (column 4) in 2015 according to their internal grades. *Mother with high education* and *Father with high education* indicate the % of 8^{th} graders in that school whose mother/father had more than a primary school degree in 2015. *Teacher-pupil-ratio* is the average number of 8^{th} grade students per class in 2015, and *Long school-day* is a dummy for whether the school had a longer instruction day in 2015. All regressions include school district fixed effects. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.6: Cross-Sectional Correlation Between Covariates and Learning Outcomes in 2015

	Math		Language	
	Average ECE	Average Internal	Average ECE	Average Internal
Spanish as native tongue (% of students)	0.672***	0.044	0.857***	0.206***
	(0.045)	(0.056)	(0.039)	(0.054)
Mother with high education (% of students)	0.959***	0.252***	1.180***	0.363***
	(0.079)	(0.089)	(0.066)	(0.092)
Father with high education (% of students)	0.474***	0.314***	0.591***	0.327***
	(0.076)	(0.084)	(0.061)	(0.083)
Repeated last year (% of students)	-0.438***	-2.578***	-0.206*	-2.772***
	(0.157)	(0.168)	(0.117)	(0.170)
Male (% of students)	-0.132**	0.334***	0.125***	0.385***
	(0.055)	(0.065)	(0.046)	(0.065)
Teacher-pupil ratio	0.008***	-0.023***	0.008***	-0.020***
	(0.001)	(0.002)	(0.001)	(0.002)
Long school-day	0.338***	0.242***	0.214***	0.269***
	(0.024)	(0.030)	(0.018)	(0.030)
Rural	-0.048**	0.023	-0.135***	-0.001
	(0.023)	(0.028)	(0.018)	(0.028)
Observations	8010	8010	8010	8010
R^2	0.491	0.243	0.676	0.239

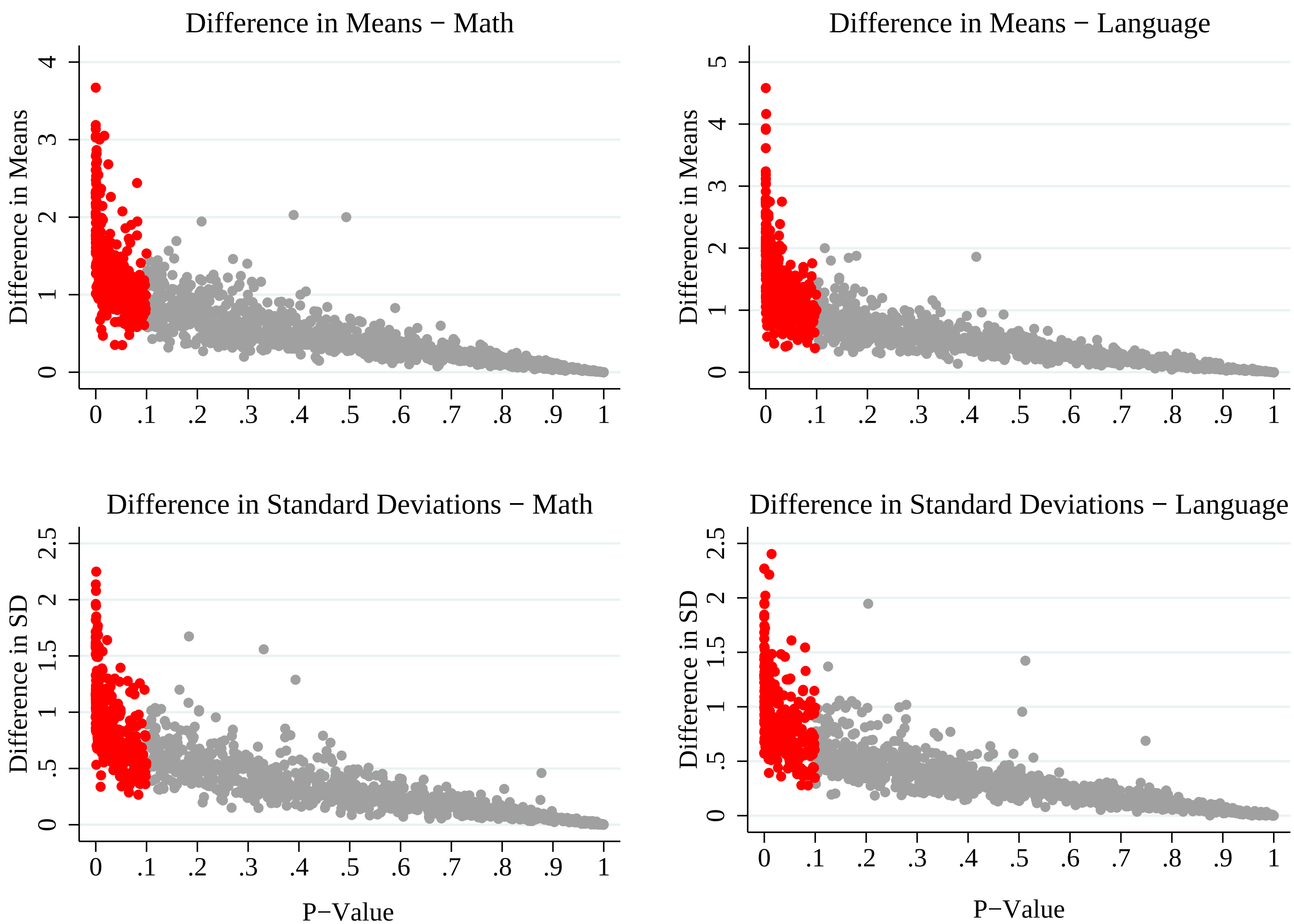
Notes: The sample includes all public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable in columns 1 and 3 are 8^{th} graders' average ECE grades in math and language, whereas the dependent variable in columns 2 and 4 are the school's average internal grade for 8^{th} grade students in 2015. Average ECE and internal grades are standardized across schools (i.e., expressed as a z-score).*Mother with high education* and *Father with high education* indicate the % of 8^{th} graders in that school whose mother/father had more than a primary school degree in 2015. *Teacher-pupil-ratio* is the average number of 8^{th} grade students per class in 2015, and *Long school-day* is a dummy for whether the school had a longer instruction day in 2015. All regressions include school district fixed effects. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.7: Within-School Correlation Between Covariates and Learning Outcomes in 2015

	Math		Language	
	Grade (z-score)	Low Achievement	Grade (z-score)	Low Achievement
Panel A: ECE Grades				
Socioeconomic status index	0.135***	-0.044***	0.190***	-0.049***
	(0.003)	(0.001)	(0.003)	(0.001)
Male	0.220***	-0.071***	0.020***	-0.010***
	(0.004)	(0.002)	(0.004)	(0.002)
Observations	354429	354547	354529	354547
R^2	0.020	0.216	0.015	0.254
Panel B: Internal Grades				
Spanish as native tongue	0.204***	-0.016***	0.218***	-0.011***
	(0.012)	(0.004)	(0.012)	(0.003)
Mother with high education	0.158***	-0.021***	0.166***	-0.014***
	(0.005)	(0.002)	(0.005)	(0.001)
Father with high education	0.129***	-0.017***	0.136***	-0.014***
	(0.005)	(0.002)	(0.005)	(0.001)
Male	0.138***	-0.035***	0.353***	-0.051***
	(0.005)	(0.002)	(0.005)	(0.001)
Observations	324696	325320	324689	325320
R^2	0.019	0.103	0.044	0.099

Notes: Panel A contains the anonymized sample of 8^{th} graders taking the ECE standardized test in 2015, and the sample from Panel B includes all 8^{th} graders in 2015 from public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable in columns 1 and 3 of Panel A (Panel B) is the students' ECE (internal) grade in math and language, standardized by subject and school. The dependent variable in columns 2 and 4 of Panel A (Panel B) is a dummy for whether the student scored in the lowest category in the ECE (failed according to his internal grades). *Socioeconomic status* is an index ranging between -3.5 and 9.5, which is increasing in measures of socioeconomic status such as parents' education, and household assets and characteristics. All regressions include school fixed effects. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Figure 4: Variation in Internal Grades Across 8^{th} Grade Classes in 2014



Notes: Top figures depict the difference in means and the corresponding p-value when testing whether the mean math and language internal grades differ across 8^{th} grade classes from the same school in 2014, in every school with only two 8^{th} grade classes. The bottom figures depict the difference in the standard deviation and the corresponding p-value for an F-test of difference in variances in the same sample of schools.