Essays on Economics of Education and Public Policy

by

Gonzalo Andrés Gaete Romeo

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Economics

September 2017

THE UNIVERSITY OF WARWICK

Contents

List of Figures

List of Tables

Acknowledgements

I am profoundly indebted to my supervisors Clément de Chaisemartin, Roland Rathelot and Fabian Waldinger, for their complete support, useful guidance, understanding and encouragement throughout all the years of my PhD.

I am also grateful to Sascha Beker for his useful advice throughout the Programme.

I would like to thank the Warwick Economics Department, both the faculty, the administrative staff and my fellow PhD colleagues.

I want to thank the Chilean Government for funding my MSc and PhD studies in the UK.

I thank the Pérez-Bravo family for allowing my kids to grow surrounded by cousins.

I thank my parents, Margarita and Gilberto, for their unconditional love and for supporting me in every decision that I have taken. They have always been behind me in every moment of my life: in through the good ones and through the not so good ones.

I specially thank my wife, Catalina, for being my partner in this journey. We made it together. Finally, I thank my sons Beltrán and Ínigo, the most important achievement of my life.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Chapter 2 was conducted in collaboration with Rocco d'Este and Giulio Trigilia.
- Chapter 3 was conducted in collaboration with Miguel Almunia.

Abstract

In Chapter 1, I study the effect of school absenteeism on secondary school students academic outcomes using the Chilean student strikes in 2011 as a source of exogenous variation. The strikes, led by university students but promptly joined by hundreds of thousands of secondary school students, triggered a significant drop in public secondary school attendance (a decline of about 15 percentage points in all four grades). Attendance returned to normal levels in 2012. Using the type of school that students attended in 2011 as an instrument for school absenteeism, I show that school absenteeism has negative effects on secondary school students' results in a post-secondary high-stakes math exam and university enrollment rates. Instrumental variables estimations suggest that a 10 percentage point decrease in attendance during secondary school is related to a 9.5 percent of a standard deviation decline in the math exam score, and a 3.2 percentage point reduction in the associated probability of university enrollment. I do not find any significant effect on the high-stakes language exam at the 5 percent level. A key finding is the persistent negative effect of school absenteeism on students' academic performance: this negative effect is present even for those students who sat the high-stakes exams three years after the strikes had ended, that is, after three years of regular schooling following the negative shock to their attendance. These results are not driven by inputs to the education production function that might have been affected by the student strikes, such as disruptiveness at the time of the high-stakes exams, school environment, teachers, class instruction, or class size.

Chapter 2 presents the first value-added (VA) estimates for doctoral teaching assistants (DTAs). We focus on the undergraduate program of the Economics Department at a UK university, where the match between students and DTAs is random. We find that a one standard deviation change in DTA quality increases students'

test scores by around 8.5 percent of a standard deviation. A novel feature of our data allows us to examine within-course dynamics in the VA estimates: These are larger for assessments taken during term-time, drop for end-of-term tests and are not statistically different from zero for final exams. The analysis suggests that the lack of persistence of the VA measures might be connected with: (i) Students' endogenous investment responses and (ii) temporal decay in teacher-related human capital. We discuss how our results can inform the broader debate on the measurement of teachers quality via the VA approach.

In Chapter 3, we study the effects of a penalty points system (PPS) introduced in Spain in 2006. We find a 20% decrease in cumulative road fatalities in the five years after the reform, compared to a synthetic control group constructed using a weighted average of other European countries. Evidence suggests that the persistent reduction in road fatalities might not only be driven by deterring risky-driving behavior, but also by taking reckless drivers out of the roads. Using estimates of the value of a statistical life, we calculate that the PPS yielded a net economic benefit of €4.6 billion ($6 billion) over this period, equivalent to 0.43% of Spain's GDP.

Chapter 1

Follow the Leader: Student Strikes, School Absenteeism and Persistent Consequences on Educational Outcomes

1.1 Introduction

School absenteeism is a major concern in the United States education system. During the academic year 2013-14, more than 6 million students skipped 15 or more days of school,[1] which represents approximately 14% of the student population or about 1 in 7 students (US Department of Education (2016)).[2] Even though there is heterogeneity in the rates at which students of different races and ethnicities experience chronic absenteeism[3], it spikes in high school for students of every race and ethnicity: nearly 20% of the high school students in the United States miss at least 10% of the school year (US Department of Education (2016)).[4] Nonetheless, concern about absenteeism is not exclusive to the United States. The *Trends in*

[1] In the United States, the academic year is about 180 days.

[2] The Obama Administration launched a variety of national initiatives to improve school attendance, including *Every Student, Every Day* and *My Brother's Keeper Success Mentor Initiative.*

[3] Chronic absenteeism is defined as missing at least 10%, or about 18 days, in a year for any reason.

[4] 12% of middle school students and 11% of elementary school students are chronically absent.

International Mathematics and Science Study (TIMSS) conducted in 2015 provides information about students' attendance in all 39 participant countries. The international average of 8^{th} grade students that skipped classes at least once every two weeks during the school year is 16% (IEA (2015)). The *2015 Programme for International Student Assessment (PISA)* report confirms this trend (OECD (2016)). The OECD countries' average of 15-year-old students that reported skipping at least a day of school in the two weeks prior to the PISA test was almost 20%.

Empirical evidence suggests that school absenteeism matters, showing that it is correlated with a variety of outcomes: It is negatively correlated with students' performance on standardized tests (IEA (2015), OECD (2016)), it is an early predictor of dropping out of school (Romero and Lee (2007), Connolly and Olson (2012)), and it is linked to juvenile delinquency (McCluskey et al. (2004)), among others. Nevertheless, causal evidence regarding the effect of school absenteeism on students' educational outcomes is scarce.

In this paper, I treat the Chilean student strikes in 2011 as a source of exogenous variation in secondary school students' class attendance. The Chilean student strikes were the largest national strikes in Chile's history. They were the result of a conflict during the first semester of 2011 between students and the Chilean authorities about changes to the education system, which escalated into massive student-led protests across the country. University students initiated the strikes, but they were promptly joined by secondary school students. This situation endured for several months, with hundreds of thousands of secondary school students skipping classes to join the protests. After extended periods of absenteeism during the 2011 school year, secondary school students resumed classes as normal at the beginning of the 2012 school year.

The students' main demand was to change the current market-oriented education system into a public education system that provides free and high-quality education at every level (Simonsen (2012)). The protest focused on the public education system, so it was the public secondary school students who experienced a large negative shock in school attendance in 2011.[5] An average attendance rate over the 2007-2010 period of about 90% for public school students in every secondary

[5] A more detailed description of the Chilean school system is presented in Section 1.2.1.

school grade dropped sharply to 71% in 2011, but then returned immediately to pre-protest levels in 2012. This drop in attendance was mainly driven by the absences during July, August and September, a period in which the average attendance rate of students in public schools fell below 55%. This highlights that public school students from every secondary grade skipped classes for long and continuous periods during this school year. In contrast, there were much less significant reductions in attendance for students enrolled in voucher and private schools.

The Chilean education system uses high-stakes testing to rank students for admission to selective universities,[6] as is the practice in many other countries. These exams cover the whole secondary education curriculum and are taken shortly after the completion of secondary school. The combination of the timing of the high-stakes exams, and the fact that the student strikes affected public school students' attendance in each of the four secondary school grades only in 2011, allows me to study the persistent effects of school absenteeism on secondary students' academic performance. I use high-quality administrative data containing individual information for all secondary school students in Chile from 2003 to 2014.

The strikes introduced large variation in school attendance across different types of schools during 2011. I conduct an instrumental variable (IV) analysis to identify the causal effect of school absenteeism on academic outcomes, using the type of school that students attended in 2011 as an instrument for school absenteeism. Instrumental variables estimates suggest that a 10 percentage points decrease in school attendance rate during secondary school is related to a 9.5% σ decline in the high-stakes math exam score and a 3.2 percentage point reduction in the probability of enrolling in a selective university. I do not find any significant effect on the high-stakes language exam at the 5 percent level.

A key finding is the persistent negative impact of school absenteeism on students' academic performance. This negative effect is present even for those students who sat the high-stakes exam three years after the strikes had ended, that is, after three years of regular schooling following the negative shock to their attendance. This highlights the persistent effect of school absenteeism: attendance during each grade

[6] A more detailed description of the Chilean post-secondary education system is presented in Section 1.2.2.

of secondary school matters in terms of academic performance. IV estimates suggest that a 10 percentage points decrease in school attendance rate in any grade during secondary school reduces the math exam score between 2.3 - 3.2% σ, while the probability of enrolling in a selective university drops between 0.8 - 1.2 percentage points. This is a sizeable effect: The previous enrollment rate of public school students in selective universities was around 9.8%, so this is a decline of between 8.2 - 12.2%.

Robustness analysis shows that results are not driven by the sorting of students across schools following the strikes. I also provide evidence that results are not driven by the sorting of students across cohorts, induced by an increase in grade repetition rates in 2011. Furthermore, factors changing at the same time as the student strikes that only affect public school students or voucher school students may be potential threats to my identification strategy. I provide evidence that the previous results are not driven by inputs to the education production function that might have been differently affected by the student strikes across public and voucher schools, such as disruptiveness at the time of the high-stakes exams, school environment, teachers, class instruction or class size.

My study is connected to several strands of the literature. Previous research has studied the effect of absenteeism on education outcomes. Most of this research uses small data sets from university undergraduate programs and focus on economics-related subjects. Moreover, these articles only study the impact of absenteeism on students' contemporaneous performance, but not the long-term consequences for academic achievement. A common finding is the negative relationship between absenteeism and students' performance in academic tests. A key challenge for identifying this causal effect is the potential omitted variables bias that may arise from non-observables correlated with both education outcomes and absenteeism. The first papers use cross-section data, controlling for proxies of students' ability and motivation, among other covariates (Romer (1993), Durden and Ellis (1995)). Their evidence also suggests that excessive absenteeism is what really matters. Stanca (2006) and Martins and Walker (2006) use panel data to account for unobserved student heterogeneity, obtaining qualitatively similar results.

Some papers exploit exogenous variation in students' absenteeism in the context of university undergraduate education. Chen and Lin (2008) conducted an experiment in which some course material was randomly skipped in different sections of the same course, and students all sat the same exam at the end of the semester. Their findings suggest that attending lectures corresponds to a 9.4% to 18.0% improvement in exam performance. Dobkin et al. (2010) obtained qualitatively similar results, using random variation generated by a policy for lower-scoring students, which forces them to attend classes. Arulampalam et al. (2012) exploit variation from the random allocation of students in the same course to different classes, given that absenteeism was more prevalent among students allocated to the early morning classes. However, skipping classes at university might have different effects compared to secondary school.

Few papers investigate the causal effect of absenteeism on school students' academic attainments, mainly focusing on the contemporaneous effect of absenteeism on primary school students. A general finding is that school absenteeism is more relevant for students' performance in math tests than language or reading tests, which is also revealed in my results. Gottfried (2010) implements an instrumental variable strategy, in which the distance that students live from school is used as an instrument for absenteeism. Goodman (2014) uses snowfall variation in Massachusetts as an instrument for identifying the effect of the time spent at school on achievement test scores. Aucejo and Romano (2016) jointly estimate the effect of absences and length of the school year on test results of primary public school students in North Carolina, by exploiting a state policy that varies the number of school days prior the tests. The authors also use flu data at the county level to instrument for school absences.

My study also connects with research that has used student strikes as a source of exogenous variation. Maurin and McNally (2008) studied the 1968 student riots in France, establishing exogenous variation that increased the likelihood of spending a greater number of years in higher education. Because of the conflict between students and university authorities, normal examination procedures were abandoned during that year, considerably increasing the pass rate for several qualifications. Using date of birth as an instrument, the authors find that additional years of higher education increase future wages and occupational levels. They also find a transmission of the effect across generations, reflected in children's educational attainment.

González (2016) uses the Chilean student strikes in 2011 to investigate the role of networks in collective action. His main finding suggests that individual participation in the strikes follows a threshold model of collective behavior: students were influenced by their networks to skip classes only when more than 40% of their network's members also skipped classes. González (2016) also investigates the effect of the student strikes on some students' academic achievement. In particular, he studies the impact of the strikes on GPA and repetition rates, by comparing students in primary and secondary school. His findings suggest that repetition rates increased by around 3.5 percentage points and that GPA decreased by 0.1 σ in 2011 among secondary school students. My research differs from Gonzalez's work in several respects: First, by using different standardized tests I am able to analyze the effect on different areas of learning (math and language). Second, in contrast to school GPA, I use national-level standardized tests which take place at an external location and are graded by an external institution.[7] Thirdly, I also study the effect on the students' transition to post-secondary education.[8]

The remainder of the paper is organized as follows. Section 1.2 describes the

[7]A more detailed description of the high-stakes exam testing system is presented in Section 1.2.2.

[8]My paper connects with previous research that has studied the effect of different types of shocks to instruction time on students' academic performance. Belot and Webbink (2010) and Baker (2013) study the effects of teacher strikes on students' educational outcomes. Herrmann and Rockoff (2012) and Duflo et al. (2012) study the impact of teacher absences on students' academic attainments. Lavy (2015) exploits cross-country variation in weekly hours of instructional time per subject to study the effects on students' achievement using data from PISA 2006 tests. Pischke (2007) studies the effect of drastically shortening the school year while keeping the education curriculum fixed. His findings suggest an increase in grade repetition in primary school and fewer students attending higher secondary school tracks among students in schools with a short school year. My work is also linked to studies of transitory shocks that disrupt the accumulation of human capital for secondary school students during extended periods. Aizer and Doyle (2015) analyze the effect of juvenile incarceration on crime and the formation of social and human capital among a population of juvenile offenders in Chicago, using the incarceration tendency of randomly assigned judges. They show that incarceration decreases high school graduation by 13 percentage points. The strongest results are for those aged 15 and 16, a similar age to the students in my sample.

Chilean education system. Section 1.3 provides background to the student strikes in 2011. Section 1.4 describes the data sources and Section 1.5 lays out the identification strategy. Results addressing the impact of the student strikes are discussed in Section 1.6. Section 1.7 discusses the effect of school absenteeism on students' academic outcomes, using the student strikes as a source of exogenous variation on students' school attendance. Section 1.8 concludes.

1.2 Overview of the Chilean Education System

1.2.1 Primary and secondary education

The Chilean school system is divided into eight years of primary school (1^{st} grade to 8^{th} grade) and four years of secondary school (9^{th} grade to 12^{th} grade). There are three types of schools: public, voucher, and private. Public schools are publicly administered and free of charge, voucher schools are privately owned and receive public funding per student enrolled via a voucher system, and private schools are privately owned and do not receive public funding. In 2010, these schools accounted for 42.1%, 50.8% and 7.1% of student enrollment respectively (MINEDUC (2015)). The current system evolved out of a reform passed in 1980, changing the school funding system by introducing a voucher per student. This voucher is directly paid to public and voucher schools, based on students' attendance.

The school year is about 180 days. It starts in March and ends in December, with a two week break in July. Students in 12^{th} grade finish the school year 2 weeks before the post-secondary high-stakes exams in mid-December.

1.2.2 Post-Secondary Education

Admission to post-secondary education is partly centralized. There are two types of institutions: *selective institutions*, the most prestigious universities;[9] and *non-selective institutions*, which includes some universities, professional learning insti-

[9] There are 25 universities in the Council of Chancellors of Chilean Universities (*Consejo de Rectores de las Universidades Chilenas (CRUCH)*, popularly known in Chile as *"universidades tradicionales"*). In 2010, selective institutions accounted for 57% of the university enrollment and 17% of the post-secondary enrollment.

tutes and technical training centers. To apply to a selective institution, students must take high-stakes exams after finishing secondary education, in which math and language are mandatory and there are optional exams depending on the degree. Admission is based on these high-stakes exams and students' secondary school GPA. These exams are called *Prueba de Selección Universitaria (PSU)* and they cover the whole secondary education curriculum. The examinations take place simultaneously across the country shortly after the end of the school year. These exams take place only once a year, and can be taken multiple times, for a fee.[10] The high-stakes exams contain 80 multiple choice questions and are marked electronically by an external institution.

The post-secondary education system has been widely debated. It has been criticized for excessive tuition fees, quality issues related to its rapid expansion and serious problems of information asymmetry, among other issues (Reyes et al. (2013)). Chile currently has the second most expensive private university system of any OECD country, after the United States. It is estimated that Chilean families pay directly more than 75% of the costs associated with higher education, compared to 40% in the United States and just 5% on average in Scandinavian countries. This was the background to the student strikes in 2011.

1.3 The Student Strikes in 2011

The Chilean student strikes was a wave of student protests across Chile in 2011, peaking between July and September. The strikes were a reaction to the market-oriented education system established in the early 80's, which has produced large profits for private supplier institutions and chronic indebtedness for thousands of post-secondary students (González and Montealegre (2012), Simonsen (2012)).

The strikes were initiated in late April 2011, when more than 6,000 students occupied a private university to protest after it was taken over by a private investment fund. On the 28^{th} of April, the association of university students *Confederación de estudiantes de Chile* (CONFECH)[11] convened the first protest of the year, which

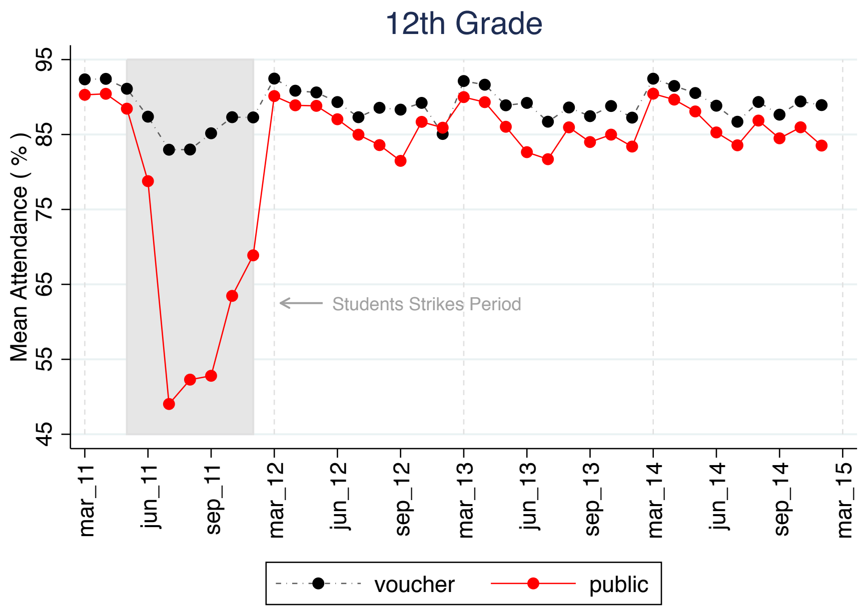
[10] The fee in 2016 is CLP 28.790 (about US$43.5) and is waived for students from public and voucher schools who apply for this benefit.

[11] This confederation brings together students from the Council of Chancellors of Chilean Uni-

brought more than 15,000 students to the streets of downtown Santiago. At this time, secondary school students started to raise their own demands relating to the deterioration of the public school education system. This first protest was followed by more than 75 others across the country over the school year.

On the 12^{th} of May, a new protest was again organized by the CONFECH, under the slogan *"national strike for saving public education"*. Again, thousands of students marched in the streets of Santiago. At this point, the student movement started to gain followers other than students, while the government suffered a drop in public support (see appendix Figure A.1). A primary demand of the movement emerged: a public education system that provides free and high-quality education at every level (Simonsen (2012)).

Figure 1.1: Average monthly attendance rates in 12^{th} grade



Notes: The figure plots the average monthly attendance rates in 12^{th} grade during the academic years 2011-2014.

In June, the situation climaxed: Universities were occupied, dozens of schools were on strike or shut down, and protesters flooded streets throughout the country, with more than 400,000 people demonstrating. Hundreds of thousands of secondary school students skipped classes to join the strikes. Figure 1.1 shows the monthly

versities, whose students are organized in democratically elected federations.

attendance rates of 12^{th} grade students over time.[12] Appendix Figure A.2 provides the monthly attendance rates for 9^{th}, 10^{th} and 11^{th} grade students. The pattern is the same: attendance of public secondary school students in all four grades dropped sharply during the period of the strikes, but immediately recovered by 2012, once the strikes were over. At the peak of the strikes (July, August and September of 2011), attendance of public secondary school students fell below 55%. There was hardly any drop in attendance at voucher schools, showing that public secondary school students were the active participants in the strikes.[13]

Around October the movement started to decay, as the end of the school year drew closer. The year ended without a clear agreement between the Government and the students.

1.4 Data

The data includes administrative records for individual-level secondary school enrollment, high-stakes exams test scores and university enrollment for 2003 to 2014. The first data source is the Chilean Students' Registry, which contains the complete population of students and is administered by the Ministry of Education. It provides information about basic demographics for each student, their annual average attendance, GPA[14] and the school each student was enrolled in. From 2011 onwards, it includes students' monthly average attendance.

The second source is the registry of students who enroll for the PSU test. This is census data provided by DEMRE (the Department of Educational Evaluation, Measurement and Recording), which depends on the Council of Chancellors of Chilean Universities. The variables I use in my analysis are individual data on PSU scores,[15] and the outcomes of the applications for post-secondary placement. The third data

[12] Public and voucher schools report the students' daily attendance to the Ministry of Education monthly, and schools receive public funds based on this information. To ensure the veracity of these reports, there is a permanent audit process.

[13] Appendix Figure A.3 provides the school attendance's distribution at student and school level during the academic year 2011.

[14] GPA is a continuous variable that goes from 1 to 7.

[15] PSU scores are normalized to a distribution with mean of 500 and a standard deviation of 110 to enable comparison between years. The scores range from 150 to 850 points.

source is the Chilean Teachers' Registry, which contains the complete population of teachers in the school system. This data is administered by the Ministry of Education and provides basic demographics for each teacher and their qualifications. The last source corresponds to teachers' 8^{th} grade SIMCE questionnaires,[16] that provides information about a variety of school environment measures for the years 2009, 2011, 2013 and 2014. This data is administrated by the Quality Assurance Agency for Education

The data is merged using individual national identification numbers provided for students, teachers and schools.

Table 1.1 shows descriptive statistics for 12^{th} grade students nationwide from 2007 to 2010, organized by type of school administration. Public school students have a slightly lower average attendance rate, a lower average take-up rate of high-stakes exams, perform worse in those exams, and fewer of them enroll in selective universities.

Over the pre-strike period, private schools account for 8.14%, voucher schools account for 45.26%, and public schools account for 46.60% of enrollment of 12^{th} grade students.

1.5 Identification

1.5.1 Estimating the effect of school absenteeism

Using this data set, I estimate the effect of school absenteeism on secondary school students' academic outcomes with the following regression model:

$$\begin{aligned} Outcome_{is_{2011}r_{2011}t} = \alpha_{s_{2011}} + \gamma_{r_{2011}t} + \rho Absenteeism_{is_{2011}r_{2011}t} + \\ \mathrm{X}'_{is_{2011}r_{2011}t}\varphi + \epsilon_{is_{2011}r_{2011}t} \end{aligned} \tag{1.1}$$

[16]SIMCE (*Sistema de Medición de la Calidad de la Educación*) is a battery of standardized tests taken some years in specific grades and it is used to measure certain aspects of the school curriculum. In addition to the tests, questionnaires are provided to students, parents and teachers, to collect information regarding specific subjects.

Table 1.1: Descriptive statistics for 12^{th} grade students over the period 2007-2010

	Private		Voucher		Public	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Students Characteristics						
Female (%)	49.10		52.90		51.92	
Yearly attendance rate	92.99	6.91	91.82	7.78	89.97	8.62
12^{th} grade GPA	6.01	0.52	5.53	0.56	5.48	0.56
Repetition rate (%)	0.27		1.87		2.68	
PSU: take-up rate (%)	97.45		76.77		64.16	
PSU: math test score (standardized)	1.26	0.82	0.08	0.89	-0.23	0.90
PSU: language test score (standardized)	1.17	0.81	0.10	0.89	-0.24	0.93
Enrollment rate in selective universities (%)	35.63		15.28		9.73	
Size of the cohort	16,549.46	124.5	92,025.25	1,570.67	94,735.75	2,729.92
Schools Characteristics						
Class Size	24.58	7.32	31.88	10.30	31.60	11.18
Rural (%)	2.91		6.38		9.96	

Here the subscript i refers to student, s refers to the school, r refers to the region, and t refers to the year. Outcomes of interest are students' high-stakes exams scores in math and language, and university enrollment. Students' high-stakes exams results are standardized within each year and university enrollment is measured with an indicator variable that takes the value of 1 if the student was enrolled in a selective university right after school graduation.[17]. I focus my regression analysis on two types of students: public secondary school students and voucher secondary school students.[18]

In Equation (1.1), I regress students' academic outcomes on students' school absenteeism rate during secondary school. The analysis focuses on ρ, which indicates the average effect of school absenteeism on students' outcomes. I control for measures of students' ability using their rank position in their class 4 years before sitting the high-stakes exams.[19] As the last cohort of 12^{th} grade students I use in the analysis is the 2014 cohort, this measure was fixed before the strikes for all cohorts included in the analysis. The other student-level control used is a gender dummy. I also control for time-invariant unobserved heterogeneity across schools by including school fixed effects. In addition, to account for potential heterogeneous regional responses to temporal shocks I include region $\times$ year fixed effects.

1.5.2 Using the student strikes as instrument for school absenteeism

Estimating Equation (1.1) using ordinary least squares (OLS) would lead to a biased estimate of ρ. Omitted variables could be a source of bias. For instance, more able and motivated students are more likely to attend school and to perform well on tests.

[17]As discussed in Section 1.2.2, students are allowed to sit high-stakes exams more than once, and to apply to a major degree as many times as they wish.

[18]I decided to exclude private secondary school students from the sample for two reasons: (i) Private school students represent a very small proportion of the total of secondary school students and; (ii) family backgrounds of private school students are very different from the rest of the student population. Their families can afford very expensive fees for private education. Thus, strikes might have differently affected their unobserved characteristics in comparison to the rest of secondary school students.

[19]This measure of students' past performance is a continuous variable that goes from 0 to 100.

To tackle this problem, I use the type of school that students were attending in 2011 to instrument students' school absenteeism during secondary school, taking advantage of the large variation introduced by the student strikes in school attendance across different types of schools during 2011.

Appendix Figure A.5 shows the students' attendance rates during secondary school, displaying a sharp decline public school students' attendance rates immediately after the strikes which remains throughout the post-strike period. This is because school attendance for all post-strike cohorts of 12^{th} grade students was similarly affected by the strikes in 2011, even though in 2011 they were enrolled in different grades of secondary education.

The first-stage regression is:

$$\begin{aligned} Absenteeism_{is_{2011}r_{2011}t} = \alpha_{s_{2011}} + \gamma_{r_{2011}t} + \phi(Public_s \times 2011_t) + \\ \mathrm{X}'_{is_{2011}r_{2011}t}\varphi + \varsigma_{is_{2011}r_{2011}t} \end{aligned} \tag{1.2}$$

In Equation (1.2), the instrument for school absenteeism during secondary school is $(Public_s \times 2011_t)$, which is a dummy variable that takes the value of 1 if the student attended a public school in 2011, 0 otherwise.

For each student in cohorts after 2011, I assign the school (and therefore also the region) attended in 2011. For these students I decided to fix the school they attended in the year of the strikes, given that the choice of the school on which they graduate from 12^{th} grade is endogenous. As the effect of the student strikes is likely to be correlated within schools, I account for any dependence between observations of students within the same school by clustering all regression results at the school level.

Using the student strikes as an instrumental variable relies on the assumption that the student strikes only affected secondary school students' academic outcomes through the effect on students' school absenteeism. It is worth mentioning that any factor affecting public and voucher secondary school students in the same way will be captured by the year fixed effects and would thus not invalidate the identification strategy. Thus, only factors changing at the same time as the student strikes that only affect public school students or voucher school students may be potential threats to the identification strategy. I will discuss potential factors in Section 1.7.3.

1.6 The Effect of the Student Strikes

I begin the empirical analysis by studying the effect of the strikes on students' school attendance. Figure 1.1 and appendix Figure A.2 show that the drop in the monthly attendance rate of public secondary school students in all four grades is very large in comparison to voucher secondary school students during the strikes, but immediately recovered in 2012 when the strikes had ended. At the peak of the strikes (July, August and September), the relative fall in monthly attendance was more than 30 percentage points.

However, high-stakes exams and university enrollment happen once a year, at the end of the school year. Appendix Figure A.4 presents the yearly attendance rates of secondary school students by grade. Again, attendance rates follow a similar pattern across the different grades of secondary school. The average yearly attendance rate of students in public secondary schools dropped sharply in 2011, but immediately recovered in 2012. To estimate the effect of the student strikes on the yearly attendance of public secondary school students compared to voucher secondary school students, I regress students' yearly attendance on the interaction between 2011_t, an indicator variable that takes the value of 1 in year 2011, and $Public_s$, another indicator variable that is 1 if the student attends a public school. I control for school fixed effects, region $\times$ year fixed effects, and the student-level controls.

$$YearlyAttendance_{isrt} = \alpha_s + \gamma_{rt} + \varphi(2011_t \times Public_s) + \mathrm{X}^{'}_{irst}\varphi + \upsilon_{isrt} \tag{1.3}$$

Regression results of Equation (1.3) are presented in Table 1.2. Column 1 shows the effect of the student strikes on the yearly attendance rate of 9^{th} grade public school students, Column 2 shows the results for 10^{th} grade public school students, Column 3 shows the results for 11^{th} grade public school students, and Column 4 shows the results for 12^{th} grade public school students. Estimates reveal a significant reduction in the yearly attendance rate of public secondary school students in all

Table 1.2: Effect of the student strikes on public school students' yearly attendance in 2011

	(1) 9^{th} Grade	(2) 10^{th} Grade	(3) 11^{th} Grade	(4) 12^{th} Grade
public $\times$ 2011	-14.87***	-14.91***	-15.35***	-14.34***
	(0.888)	(0.914)	(0.957)	(0.899)
Observations	1,965,085	1,774,719	1,613,636	1,436,384
R-squared	0.311	0.305	0.304	0.301
Student Level Controls	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Region $\times$ Year FE	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are yearly attendance rates of students in 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. These variables are measured in percentage points and goes from 0 to 100. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time $\times$ region fixed effects.

grades compared to students in voucher schools. This average reduction is very similar across grades, with point estimates between 14.3 - 15.3 percentage points. These results imply that on average different cohorts of students in each of the four grades of secondary education were exposed to the same large and negative transitory shock to their attendance in 2011.

Next I study the effect of the student strikes on public secondary school students' academic outcomes, using the following reduced form regression:

$$\begin{aligned} Outcome_{isrt} = \alpha_s + \gamma_{rt} + \sum_{m=2007}^{2010} \beta_m (Year_t \times Public_s) + \\ \sum_{n=2011}^{2014} \varrho_n (Year_t \times Public_s) + x'_{irst}\varphi + \epsilon_{isrt} \end{aligned} \tag{1.4}$$

This regression is a differences-in-differences estimate of the effect of the student strikes on public secondary school students' academic outcomes. In Equation (1.4)

I also control for school fixed effects, region $\times$ year fixed effects, and the student-level controls. The key assumption in a differences-in-differences strategy is that the evolution of outcomes in public and voucher schools would follow the same trend. If the common trend assumption is not satisfied, it is expected that coefficients β_m are statistically different from zero. In addition, Equation (1.4) offers a more flexible estimation of the strikes' impact on secondary school students' outcomes. Indeed, it decomposes the effect, making it possible to study its evolution over time. These yearly effects are captured by coefficients ϱ_n.

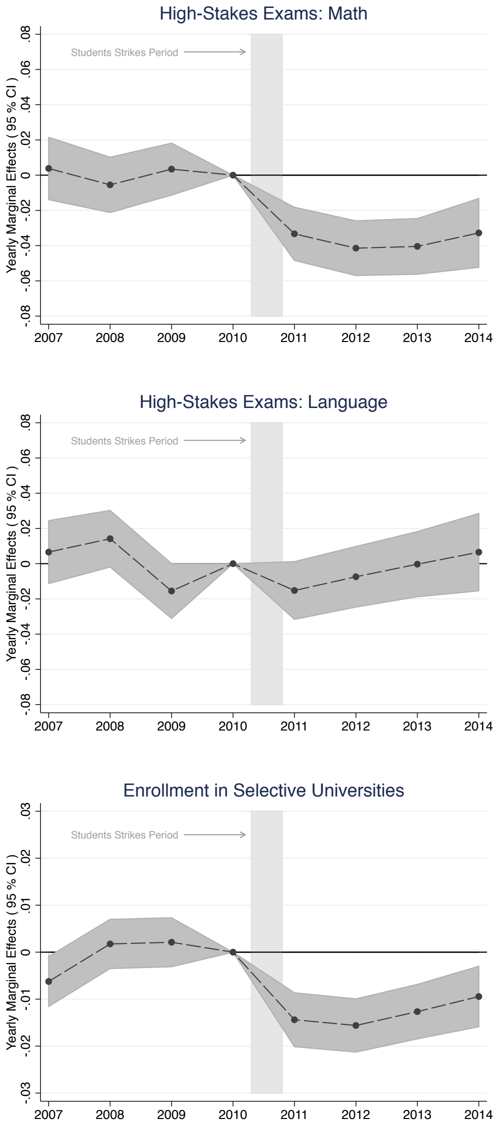
Figure 1.2 graphs the coefficients of Equation (1.4) separately for each year and appendix Table A.1 shows the regression results.[20] At first glance, it is possible to highlight two features: (i) In all outcomes, coefficients β_m are not statistically different from zero; and (ii) in math test and university enrollment, there is a negative effect of the student strikes on public school students that persists for the whole post-strike period, even though students resumed classes as normal in 2012. This second feature is interesting. It is reasonable to expect that the strikes could have an immediate negative effect on students' performance, however, the impact lasts well beyond 2011. Public school students who were in their penultimate year of secondary education at the time of the strikes were negatively affected one year later, when they sat their post-secondary high-stakes examinations in 2012. Similarly, students who were in 10^{th} and 9^{th} grade in 2011 were negatively affected in their post-secondary exams, 2 and 3 years after the strikes.

In the high-stakes math exam, the effect on the four cohorts of students that sat the exams after the strikes is relatively constant and around 4.0% σ. The immediate impact of the strikes on university enrollment rates is 1.4 percentage points and remains significant at the one percent level for the post-strike period. By contrast, I do not find any significant effect on the high-stakes language exam at the 5 percent level. The effect on the enrollment rate in selective universities has a similar pattern to that on the high-stakes exams, because university enrollment depends on the exam results.

A potential driver of the education outcomes after 2011 may have been that

[20] Appendix Figure A.6 presents in raw data the evolution of the three main academic outcomes of interest in both types of schools. To make the comparison easier, levels in the year 2007 are set to 0.

Figure 1.2: Effect of the student strikes on 12^{th} grade public school students' academic outcomes



Notes: Top panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes math exam of 12^{th} grade students as a dependent variable. Middle panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes language exam of 12^{th} grade students as a dependent variable. Bottom panel plots parameter estimates of Equation (1.4), using the enrollment status in a selective university of 12^{th} grade students right after finishing secondary school as a dependent variable. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2014. I include school fixed effects and region $\times$ time fixed effects. At student level, I control for pre-strikes measures of individual students' past performance. I also include a gender dummy.

good students moved from public schools more affected by the strikes to less affected voucher schools. I address this by assigning students who switched schools in 2012, 2013 and 2014 to their 2011 school. Appendix Figure A.7 shows the new results, which are qualitatively the same as those in the main specification presented in Figure 1.2. Point estimates barely change.

Another concern is that repetition rates increased in public schools in 2011 because of the sharp decline in attendance rates (top and middle panels in appendix Figure A.8).[21] Accordingly, the take-up rates of the high-stakes exams in public schools went down in 2011 (bottom-left panel in appendix Figure A.8). As a result the 2010 and 2011 cohorts taking the high-stakes exam in public schools may not be comparable, invalidating my differences-in-differences estimation strategy for the high-stakes exam results in 2011. Moreover, this situation could have induced self-selection of students across cohorts, which also threatens the identification of my reduced form analysis. Self-selection might occur, for example, if low-achieving students were more likely to repeat a grade due to school absenteeism in 2011 than high-achieving students. I partly addressed this problem in my specification by controlling for pre-strikes measures of students' ability. Top-left and top-right panels of appendix Figure A.9 provide the measure of students' past performance over time for non-repeaters and students that took the high-stakes exams, showing that the strikes didn't appear to affect the trends of these measures. This first descriptive evidence suggests that the strikes did not impact the composition of students' ability across cohorts. Furthermore, I conduct two robustness checks: (i) I put students that repeated after 2011 back in their original cohorts; and (ii) I use an estimation strategy inspired on Abadie (2003)[22] to characterize the students that were *induced* to repeat by the strikes in terms of their academic past performance. Appendix Figure A.10 shows the results of the first exercise, which are qualitatively the same as those obtained in the main specification presented in Figure 1.2. The result of the second exercise is provided in Appendix **??**. Equation (A.4) shows that the mean ability of the students that repeated a grade *induced* by the strikes in 2011

[21] In Chile, students' attendance rates can be a reason to repeat the academic year if the annual attendance rate of a student is less than 85%; the school principal can consider individual cases.

[22] Abadie (2003) proposes an estimation method to describe compliers in instrumental variable models.

is similar to the mean ability of non-repeaters, allaying concern about self-selection across cohorts in terms of students' ability.

1.7 The Effect of School Absenteeism on Academic Outcomes

1.7.1 Aggregate School Absenteeism during Secondary School

I use the student strikes as an exogenous shock to estimate the effect of school absenteeism on secondary students' academic outcomes. The high-stakes exams cover the whole secondary education curriculum, so school attendance for all years of secondary school could affect exam results..

Table 1.3 has the results of the first-stage regressions in Columns 1 and 6, showing a strong and significant effect of the student strikes on absenteeism during secondary school.[23] This was expected from trends shown in appendix Figure A.5. Results regarding academic outcomes are presented in Columns 2 to 5 , 7 and 8. Columns 2, 4 and 7 report the OLS results in math, language and enrollment in a selective university. Columns 3, 5 and 8 contain the IV estimates. Three features stand out: (i) OLS and IV point estimates show a negative effect of absenteeism during secondary school on all academic outcomes; (ii) OLS and IV estimates show a larger effect in math than in language; and (iii) OLS and IV point estimates are similar in magnitude for the impact of absenteeism on math and university enrollment. Columns 2 and 3 show the effect of absenteeism on the high-stakes math exam. The IV estimate suggests that a 10 percentage points[24] decrease in school attendance rate during secondary school is related to 9.5% σ reduction in math exam score. Students' performance in the high-stakes language exam is shown in Columns 4 and 5. The IV

[23] First-stage regression reported in Column 1 only contains 12^{th} grade students that sat the high-stake exams, while the first-stage regression reported in Column 6 includes the whole population of 12^{th} grade students.

[24] I express results in terms of a 10 percentage points decrease in attendance for two reasons: (i) Because chronic absenteeism is defined as missing at least 10% of the school year for any reason; and (ii) the average yearly absenteeism rate in 12^{th} grade public school students in Chile is about 10%.

point estimate suggests that a 10 percentage points decrease in the attendance rate during secondary school reduces the language exam score by 2.9% σ. Nevertheless, this effect is only significant at the 10 percent level. Finally, Columns 6 and 7 show the estimates of the effect on enrollment in selective universities. Again the IV estimate is negative and highly significant, with a 10 percentage points decrease in attendance rate during secondary school reducing the probability of enrolling in a selective university by 3.2 percentage points.

It is informative to compare my results with previous findings in the literature. My results show a larger effect of school absenteeism on math scores than language scores, which is a general finding in prior papers (Gottfried (2010), Goodman (2014), Aucejo and Romano (2016)). In particular, my instrumental variable estimates are very similar to the baseline results in Aucejo and Romano (2016). In this paper, the authors estimate the effect of absences and the length of the school calendar on test score performance of primary public school students in North Carolina. Their findings suggest that 10 days of primary school absence reduces math scores by 5.5% σ, and 2.9% σ in reading, under their preferred specification. This implies that a 10 percentage points decrease in the yearly attendance (18 days) would reduce math scores by 9.9% σ.

Interestingly, previous research into other types of shocks to instructional time has also found larger effects on performance in math tests than in language tests (teacher strikes (Baker (2013)), teacher absences (Herrmann and Rockoff (2012)), cross-country variation in instructional time (Lavy (2015)), among others). These findings and my results suggest that the education production functions of math and language are probably different and that the role of the teacher and the classroom learning process are more important for the former.

Table 1.3: Effect of secondary school absenteeism on academic outcomes

		Math		Language		Enrollment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Absenteeism	OLS	IV	OLS	IV	OLS	IV
public × 2011	4.041***						
	(0.232)						
absenteeism in secondary (%)		-0.00891***	-0.00952***	-0.00155***	-0.00293*	-0.00371***	-0.00318***
		(0.000338)	(0.00157)	(0.000287)	(0.00158)	(0.000141)	(0.000578)
Observations	1,429,263	1,033,404	1,033,404	1,033,404	1,033,404	1,429,263	1,429,263
R-squared	0.385	0.478	0.478	0.431	0.431	0.250	0.250
Student Level Controls	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES	YES	YES	YES
F-test on instrument	303.2						

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are school absenteesim during secondary school education, the standardized score in the high-stakes math exam, the standardized score in the high-stakes language exam and enrollment in a selective university, respectively. Column 1 reports the first-stage estimates only considering the population of students that sat the high-stakes exams, while column 6 reports the first-stage estimates considering the whole population of students. Columns 2, 4, and 7 report the OLS estimates regarding the effect of school absenteesim during secondary school education on academic outcomes, while columns 3, 5, and 8 reports the IV estimates. School absenteesim during secondary school education is measured in percantage points, and therefore it is a continuous variable that can take values from 0 to 100. High-stakes exams scores are standardized within each year. Enrollment is a dummy variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects. In addition, for each student in cohorts after 2011, I assign them the school (and therefore, also the region) she attended in 2011.

Compared to other interventions, reducing absenteeism during secondary school by 10 percentage points has a similar effect in math to a 1σ improvement in teacher value added in the context of primary education (Rothstein (2010a), Chetty et al. (2014b)). Nevertheless, in the context of first-year of post-secondary education (students closer in age to my sample), my estimates are almost double (Scott E. Carrell (2010), Braga et al. (2016)). In addition, reducing absenteeism during secondary school by 10 percentage points has more than the double the effect in math than offering large financial incentives to secondary school teachers based on their students' test performance (Lavy (2009)).

Heterogeneity

On average, absenteeism during secondary school has a negative impact on students' academic outcomes. However, these effects might differ depending on students' characteristics. I study heterogeneous responses to school absenteeism by dividing the sample into students above, and students below, the median of the distribution of past performance, which is the proxy for students' ability.

The results are in Tables 1.4 and 1.5. Table 1.4 contains the results on the high-stakes exams, with the results of the first-stage regressions in Columns 1 and 2 by students' ability. Columns 3 to 6 show IV estimates of the effect of school absenteeism on the math and language exams by students' ability. The IV point estimates for high-performing students are about double, indicating that public school students in the upper part of the ability distribution are more affected by school absenteeism. Results in Columns (4) and (6) suggest that a 10 percentage points decrease in the attendance rate of high-achieving students during secondary school reduces their math exams score by 11.9% σ and their language exam score by 4.3% σ. Both effects are significant at the 1 percent level. Table 1.5 lays out the effect on university enrollment, showing the same pattern. The IV point estimate in Column (4) suggests that a 10 percentage points decrease in attendance rate of high-achieving students during secondary school reduces by 3.12 percentage points their probability of enrolling in a selective university.[25]

[25] Appendix Figure A.11 presents academic outcomes divided by students' ability, showing that the larger impact on high-performing public school students is not driven by a specific year.

Table 1.4: Effect of secondary school absenteeism by students' previous performance (High-stakes exams)

	Absenteeism		Math		Language	
	Low-Achievers	High-Achievers	Low-Achievers	High-Achievers	Low-Achievers	High-Achievers
	(1)	(2)	(3) IV	(4) IV	(5) IV	(6) IV
public × 2011	3.741***	4.617***				
	(0.236)	(0.308)				
absenteeism in secondary (%)			-0.00653***	-0.0119***	-0.00215	-0.00430***
			(0.00195)	(0.00165)	(0.00206)	(0.00156)
Observations	462,753	570,622	462,753	570,622	462,753	570,622
R-squared	0.372	0.398	0.406	0.489	0.365	0.423
Student Level Controls	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES	YES	YES
F-test on instrument	250.8	224.7				

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are school absenteesim during secondary school education, the standardized score in the high-stakes math exam and the standardized score in the high-stakes language exam. Column 1 reports the first-stage estimates only considering the population of low-achieving students that sat the high-stakes exams, while column 2 reports the first-stage estimates only considering the population of high-achieving students that sat the high-stakes exams. Column 3 reports the IV estimates regarding the effect of school absenteesim during secondary school education on the high-stakes exams performance of low-achieving students, while columns 4 and 6 reports the IV estimates for high-achieving students. School absenteesim during secondary school education is measured in percantage points, and therefore it is a continuous variable that can take values from 0 to 100. High-stakes exams scores are standardized within each year. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects. In addition, for each student in cohorts after 2011, I assign them the school (and therefore, also the region) she attended in 2011.

Table 1.5: Effect of secondary school absenteeism by students' previous performance (Enrollment)

	Absenteeism		Enrollment	
	Low-Achievers	High-Achievers	Low-Achievers	High-Achievers
	(1)	(2)	(3) IV	(4) IV
public × 2011	3.629***	4.445***		
	(0.208)	(0.279)		
absenteeism in secondary (%)			-0.00249***	-0.00412***
			(0.000542)	(0.000721)
Observations	710,450	718,795	710,450	718,795
R-squared	0.375	0.386	0.182	0.269
Student Level Controls	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES
F-test on instrument	303.4	254		

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are school absenteesim during secondary school education and enrollment in a selective university. Column 1 reports the first-stage estimates considering the whole population of low-achieving students, while column 2 reports the first-stage estimates considering the whole population of high-achieving students. Columns 3 reports the IV estimates regarding the effect of school absenteesim during secondary school education on the probability of enrollment in a selective university of low-achieving students, while column 4 reports the IV estimates for high-achieving students. School absenteesim during secondary school education is measured in percantage points, and therefore it is a continuous variable that can take values from 0 to 100. Enrollment is a dummy variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects. In addition, for each student in cohorts after 2011, I assign them the school (and therefore, also the region) she attended in 2011.

On average, public school students perform worse on the high-stakes exams, even before the student strikes took place. Hence, my previous results imply that, within the most deteriorated part of the school system, high-achieving students are the most affected by the school absenteeism. This evidence suggests a complementarity between school attendance and the underlying ability in this group of students.

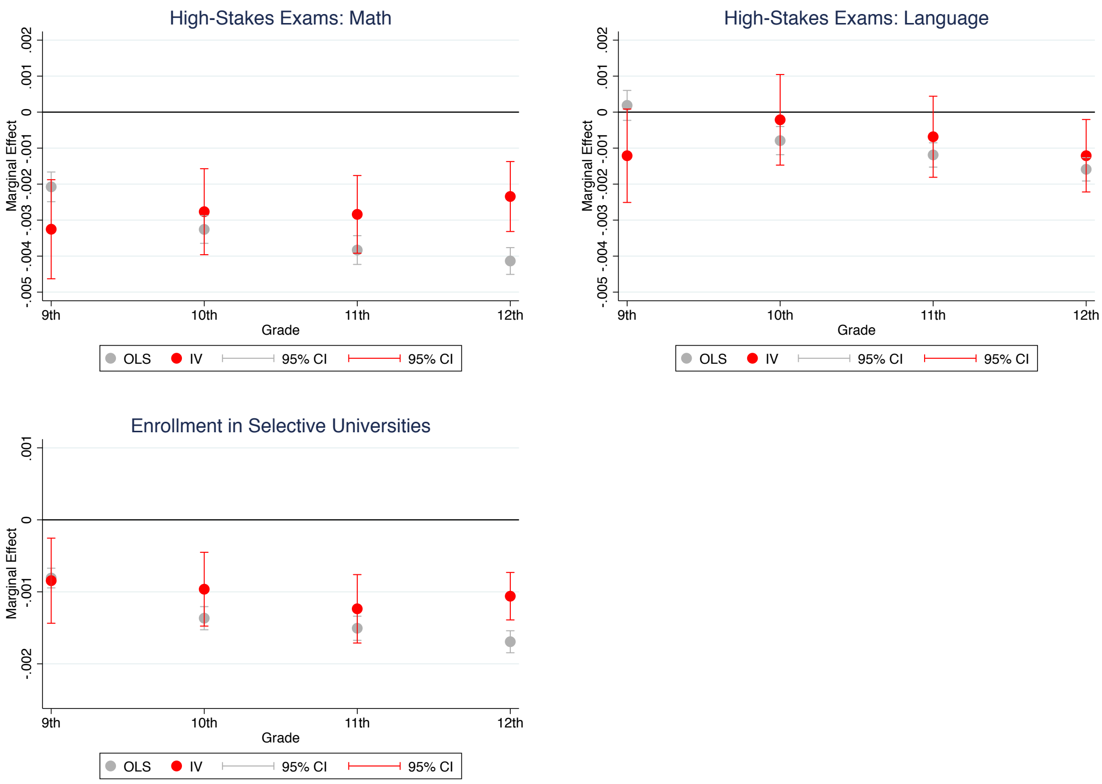
1.7.2 School Absenteeism in each Grade of Secondary School

A unique characteristic of my study is that I can estimate persistent effects of school absenteeism on students' academic outcomes. The student strikes similarly affected the attendance rate of public school students in every secondary school grade during 2011. Nevertheless, students resumed classes as normal in 2012. This means that I can study the effect of absenteeism during each grade of secondary school on high-stakes exams and enrollment in selective universities. To investigate the impact of school absenteeism during 12^{th} grade, I use Equations (1.1) and (1.2), only keeping in my sample the pre-strike and 2011 cohorts of 12^{th} grade students. The pre-strike cohorts did not receive any shock to attendance in any grade during secondary school, while the 2011 cohort was only affected in 12^{th} grade. Similarly, to study the effect of absenteeism during 11^{th} grade, I only use the pre-strike and the 2012 cohorts of 12^{th} grade students; to investigate the effect of absenteeism during 10^{th} grade, I only use the pre-strike and the 2013 cohorts of 12^{th} grade students, and to study the effect of absenteeism during 9^{th} grade, I only use the pre-strike and the 2014 cohorts of 12^{th} grade students. This is because the attendance rate of the 2012 cohort of 12^{th} grade students was only affected by the strikes when these students were enrolled in 11^{th} grade, the attendance rate of the 2013 cohort of 12^{th} grade students was only affected when these students were enrolled in 10^{th} grade, while the attendance rate of the 2014 cohort of 12^{th} grade students was only affected when these students were enrolled in 9^{th} grade.

First-stage regressions are reported in appendix Tables A.2 and A.3, showing a strong and significant effect of the student strikes on absenteeism in each grade of secondary school.[26] Results of the OLS and IV estimations regarding the effect of

[26]Point estimates in Table A.2 are slightly different from those presented in Table 1.2 for three reasons: (i) The low attendance rate of some students that repeated a grade in 2011 was replaced

Figure 1.3: Effect of school absenteeism in each secondary school's grade on students academic outcomes



Notes: Top-left panel plots OLS and IV estimates regarding the effect of school absenteeism during each secondary school's grade on the high-stakes math exam. Top-right panel plots OLS and IV estimates regarding the effect of school absenteeism during each grade of secondary school's grade on the high-stakes language exam. Bottom-left panel plots OLS and IV estimates regarding the effect of school absenteeism during each secondary's school grade on enrollment in selective universities. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2010. For estimations regarding the effect of school absenteeism in 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the cohort of 12^{th} grade students corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. High-stakes exams scores are standarized within each year, and enrollment in a selective university is a dummy variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. Attendance rate on each grade is measured in percentage points and goes from 0 to 100. The instrument for school absenteeism is $(Public_s \times 2011_t)$, which is a dummy variable that takes the value of 1 if the student was attended a public school in 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes and a gender dummy. All regressions include school fixed effects and time $\times$ region fixed effects.

school absenteeism of each grade on math, language and university enrollment are reported in Figure 1.3 and appendix Tables A.4, A.5 and A.6. A first interesting finding is the long-term implications of school absenteeism, reported in both OLS and IV estimations: attendance during each grade of secondary school matters for academic outcomes. A second interesting result is that OLS point estimates assign more importance to the later grades in students' academic performance, while IV results suggest a more even effect of each grade of secondary school. IV estimates suggest that a 10 percentage points decrease in attendance rate in any grade during secondary school reduces the math exam score by around 2.3 - 3.2% σ. In addition, IV regressions also suggest that the same decrease in attendance rate in any grade during secondary school is related to a 0.8 - 1.2 percentage points decline in the probability of enrolling in a selective university.

A potential driver of the persistent effects on math and university enrollment could be that public secondary school students were exposed to a large shock in their school attendance during the strikes. This does not make my results less interesting: Chronic absenteeism is more prevalent among teenagers, which implies that a considerable fraction of secondary school students skip a large period of the school year.

1.7.3 Robustness Checks

Using the student strikes as an instrumental variable to identify the effect of school absenteeism relies on the assumption that the student strikes only affected secondary school students' academic outcomes through the effect on attendance. Hence, factors changing at the same time as the student strikes that only affect public school students or voucher school students may be potential threats to the identification strategy. Thinking on potential violations of the exclusion restriction and on key inputs to the education production function, in this section I discuss some mechanisms that could be confounders of my results.

Transitory shocks during high-takes exams could have significant negative effects

by their new attendance rate after passing that grade later; (ii) some students dropped out of secondary school; and (iii) in the regressions reported in appendix Table A.2 I only use a subgroup of cohorts of students.

on students' exams performance (Ebenstein et al. (2016)). In 2011, students sat the high-stakes exams between the 11^{th} and the 13^{th} of December. Even though the student strikes started to decay in October, a disruptive environment on the days of the exams might affect students' academic outcomes. However, if this channel had been the main driver we wouldn't have observed an effect in 2012, 2013 and 2014.

Theoretical and empirical evidence suggests that disruption and misbehavior in the classroom affect students' outcomes (Lazear (2001), Carrell and Hoekstra (2010)). Therefore changes in the class and school environment due to the student strikes might explain my findings. To explore this alternative, I use information regarding school environment reported by teachers in the 8^{th} grade SIMCE tests in years 2009, 2011, 2013 and 2014. 8^{th} grade is the last grade of primary education, so I use information about schools that provide both primary and secondary education. This is a non-random sample of secondary education schools. In fact, this set only contains about 25% of the public schools and 56% of the voucher schools in the whole sample. Appendix Figure A.12 lays out the students' monthly attendance rate in this sub-sample of schools for each grade of secondary education, showing qualitatively the same pattern as the whole population. This suggests that the sub-sample of schools was affected by the student strikes in the same way as the whole population. The information regarding school environment is reported by specific-subject teachers (math, language, social science and natural science, which are the subjects taken in the 8^{th} grade SIMCE test), who often teach in both primary and secondary education. In particular, I have information on: (i) How difficult it is to teach in the school due to student discipline; (ii) the degree to which rules are respected by the students at the school; and (iii) the level of violence at the school. Appendix Figure A.13 presents the results of the regression analysis. There are no statistical differences across public and voucher schools in any outcome and point estimates are close to zero. This suggests that changes in the school environment are not driving the results relating to students' academic performance.

Teachers are an essential input to the education production function (Rothstein (2010a), Chetty et al. (2014b)). Thus, sorting of teachers across schools after the student strikes is a potential mechanism through which the strikes might have affected students' performance. Good teachers from more affected public schools may have moved to voucher schools less involved in the strikes. I study the turnover of secondary school teachers across schools during the period using Equation (1.4) and individual-level data of the complete population of teachers in the secondary school system. In particular, I analyze the proportion of teachers that leave the school during every year in the sample, as well as the proportion of teachers that hold an academic degree, as a measure for capturing teacher quality. Results are presented in appendix Figure A.14, showing that neither the turnover of teachers across schools nor teachers' qualifications seem to be affected by the student strikes.

Student strikes might also have affected teachers' performance across schools by making it more difficult for public school teachers to cover the curriculum. This implies that my results might not only be driven by the variation in school attendance, but also by the impact on teachers' effectiveness in public schools. Appendix Figure A.15 shows the correlation between past performance and the yearly school attendance of public secondary school students separately by year. For all years in the sample, the correlation is positive and very similar across years other than 2011. Nevertheless, this correlation is still positive but much smaller in the school year 2011. This is evidence that it was not just low-performing students who participated in the protests. Then I re-estimate Equation (1.4) separately for two groups of public school students that took the high-stakes exams on years 2011 onwards: above and below the median of their cohort school attendance in 2011. I keep all voucher school students as a control group, and I use the difference in the performance between public and voucher school students in 2010 as a reference category. Appendix Figure A.16 and A.17 shows the results. The effects on the academic attainments of the low-attendance students in 2011 are much larger in comparison to the results provided in Figure 1.2, while there is no negative effect in the high-attendance students' group. Even though students' school attendance in 2011 is an endogenous response to the strikes, this analysis allays the concern that public school students who did attend classes during the protests (who are not necessarily high-achievers in comparison to the ones protesting) were not taught the curriculum.

Finally, class size could have been affected by the student strikes, for example through a decrease in the effective class size in 2011 due to high rates of absenteeism. A second potential mechanism is a reduction in class size in public schools after

2011 if the strikes accelerated the trend of students moving from public to voucher schools.[27] Therefore, the effects I present might be attributable to a change in class size rather than the strikes. Appendix Figure A.18 shows changes in class size in each grade of secondary school education in public and voucher schools. Class size in public schools started to fall before the strikes of 2011, and continued after that. As the literature on class size suggests a positive effect of smaller classes on student achievement (Angrist and Lavy (1999), Krueger (2003), Fredriksson et al. (2012), among others), this reduction in class size should go against my results. In addition, and being aware of the potential bias induced by a *bad control problem* (Angrist and Pischke (2008)), as an indicative exercise I re-estimate my main specification using Equations (1.1) and (1.2), adding class size as a control. Results are presented in appendix Table A.7: they remain almost unchanged in comparison to the results in Table 1.3.

1.8 Concluding Remarks

I use the Chilean student strikes in 2011 to identify the effect of absenteeism on secondary school students' academic achievements. I show that student strikes, initially led by university students but spreading to secondary school students, had a very strong effect on the attendance rate of public school students in 2011. The yearly attendance rate of public school students in every grade of secondary education dropped by around 15 percentage points, compared to students in voucher schools. Nevertheless, the attendance rate of public secondary school students rebounded in 2012, when the protest action abated.

I use the exogenous variation in students' school absenteeism to estimate its effects on students' performance in post-secondary high-stakes exams and on their university enrollment. I use the type of school that students attended in 2011 to

[27]During the last 25 years, there has been a considerable flow of students from public to voucher schools. Since the introduction of the initial reform in 1980, students have moved from the public system to voucher schools. In the early 90's, 60% of the students were enrolled in public schools, 33% attended voucher schools and the remaining 7% attended private schools (Simonsen (2012)). In 2012, only 39.7% of the students were enrolled in public schools, 53.1%. attended voucher schools and the remaining 7.2% attended private schools.(MINEDUC (2015)).

instrument students' school absenteeism. Instrumental variables estimations suggest that a 10 percentage points decrease in the attendance rate during secondary school leads to a 9.5% σ reduction in score in high-stakes math exams and a 3.2 percentage point reduction in the probability of being enrolled in a selective university. In contrast, I do not find any significant effect on the high-stakes language exam at the 5 percent level.

A key finding is the persistent negative impact of school absenteeism on students' academic outcomes. This negative effect is present even for those students who had three years of regular schooling between the period of absenteeism and sitting the high-stakes exam. This highlights the persistent consequences of school absenteeism: attendance during each grade of secondary school matters in terms of students' academic performance. Instrumental variables estimates suggest that a 10 percentage points decrease in the attendance rate in any secondary school grade leads to a reduction in the high-stakes math exam score between 2.3 - 3.2% σ, while the probability of enrolling in a selective university decreases by between 0.8 - 1.2 percentage points.

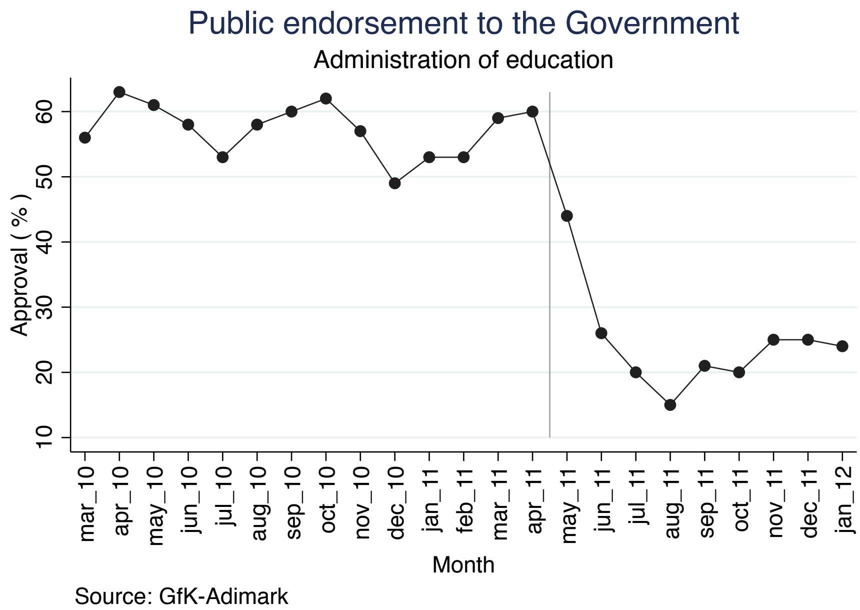
Robustness analysis shows that these findings are driven neither by the sorting of students across schools following the strikes nor by the sorting of students across cohorts induced by an increase in grade repetition rates in 2011. Moreover, factors changing at the same time as the students strikes that only affect public school students or voucher school students may be potential threats to my identification strategy. I provide evidence that the results are not driven by inputs to the education production function that might have been affected by the student strikes, such as disruptiveness at the time of the high-stakes exams, school environment, teachers, class instruction or class size.

From a policy perspective, understanding the effect of school absenteeism on secondary school students matters. Chronic absenteeism is more prevalent among teenagers and a considerable fraction of secondary school students skip a large proportion of the school year. School absenteeism could have short- and long-term impacts on students' academic achievements and merits attention from policy-makers. Furthermore, reducing absenteeism could be a cost-effective instrument for increasing students' instruction time, by making better use of the resources already allocated to schools such as classroom capacity, teacher allocation and timetabling.

Appendices

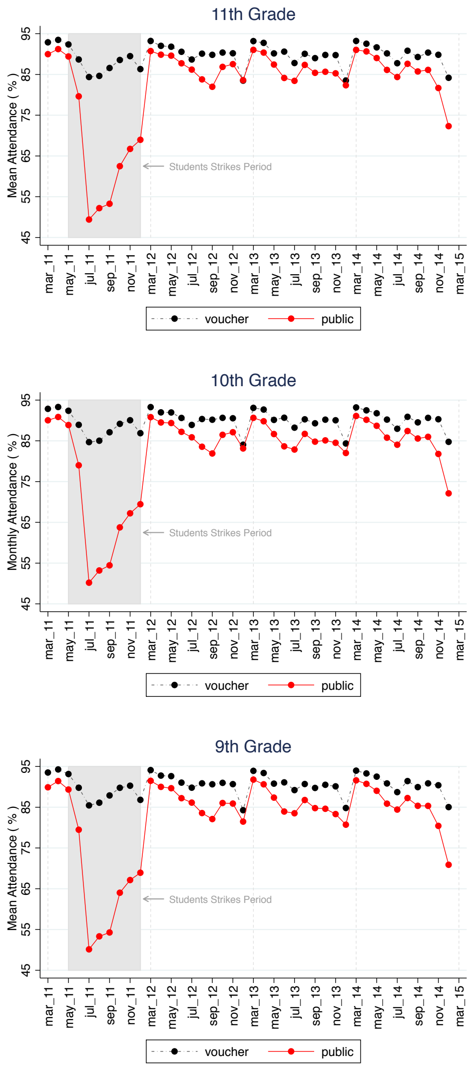
Figures

Figure A.1: Public support for Government's education policy



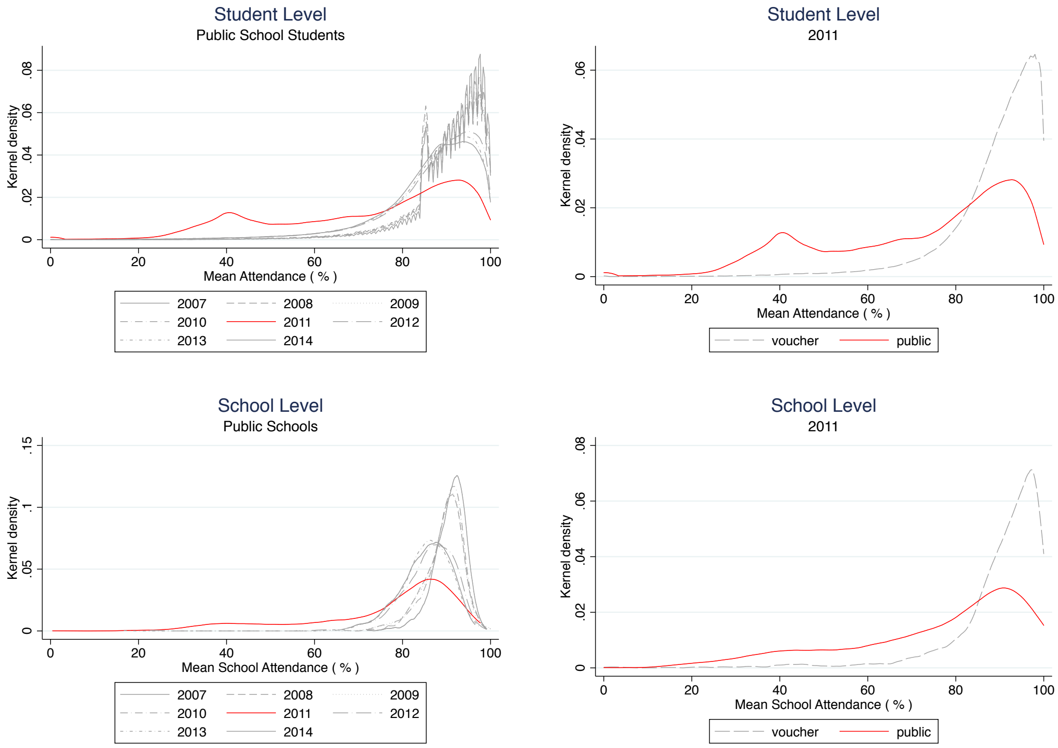
Notes: This information is captured in a monthly survey by GfK - Adimark, one of the largest Chilean firms dedicated to collect public perceptions.

Figure A.2: Average monthly attendance rates in grades 9^{th} to 11^{th}



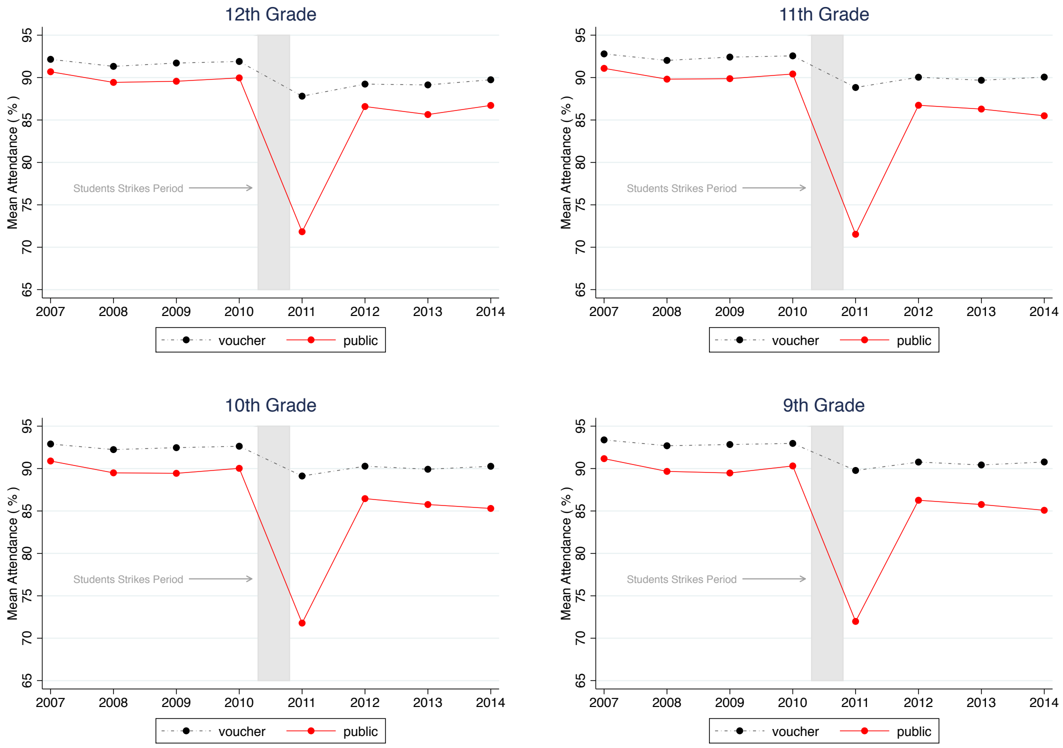
Notes: Top panel plots the average monthly attendance rates in 11^{th} grade during the academic years 2011-2014. Middle panel plots the average monthly attendance rates in 10^{th} grade during the academic years 2011-2014. Bottom panel plots the average monthly attendance rates in 9^{th} grade during the academic years 2011-2014.

Figure A.3: Distribution of School Attendance in Secondary School Education



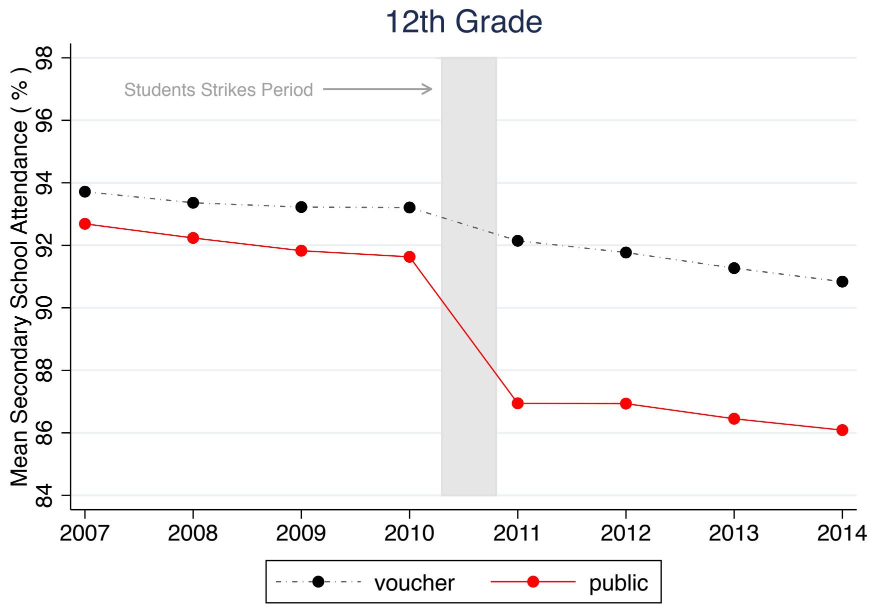
Notes: Top-left panel plots the school attedance's distribution of pubic secondary school students during the academic years 2007-2014. Top-right panel plots the school attedance's distribution of pubic and voucher secondary school students in the academic year 2011. Bottom-left panel plots the mean school attedance's distribution of pubic secondary school students at school level during the academic years 2007-2014. Bottom-right panel plots the mean school attedance's distribution of pubic and voucher secondary school students at school level in the academic year 2011.

Figure A.4: Average yearly attendance rates in grades 9^{th} to 12^{th}



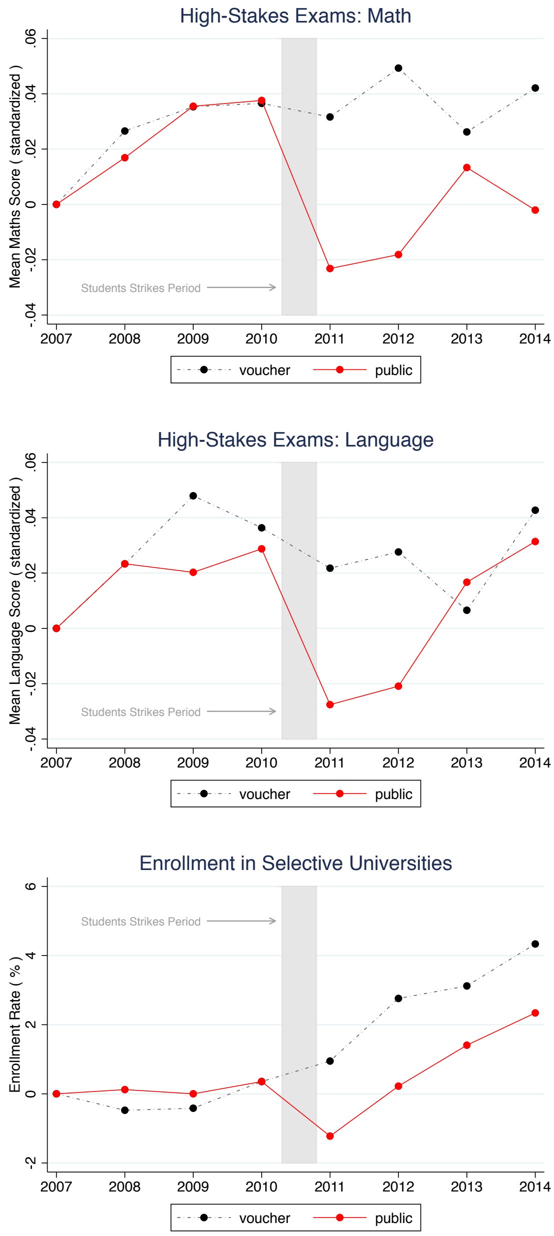
Notes: Top panel plots the average yearly attendance rates in 11^{th} grade during the academic years 2007-2014. Middle panel plots the average yearly attendance rates in 10^{th} grade during the academic years 2007-2014. Bottom panel plots the average yearly attendance rates in 9^{th} grade during the academic years 2007-2014.

Figure A.5: Secondary school attendance rates of 12^{th} grade students



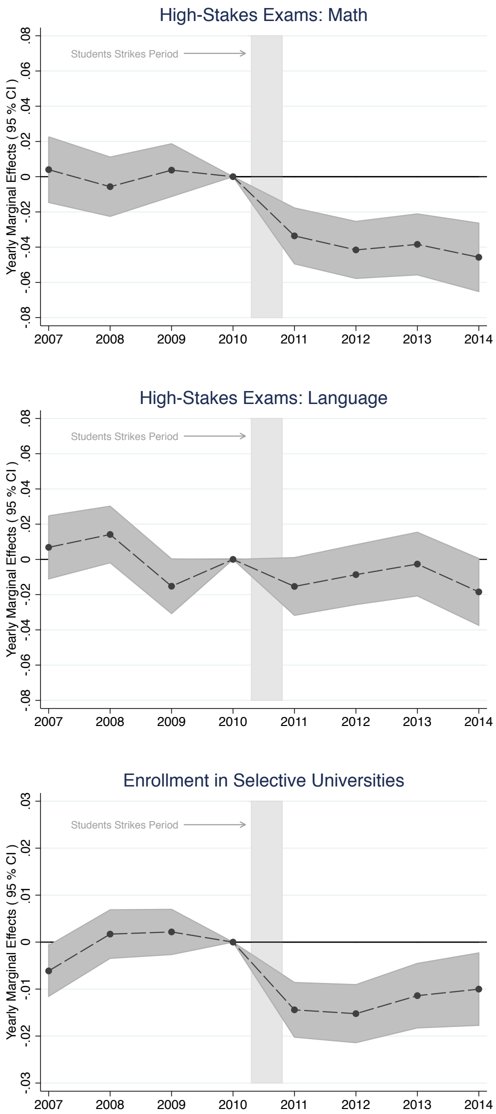
Notes: The figure plots the average attendance rate during the whole secondary school of 12^{th} grade students over the academic years 2007-2014.

Figure A.6: Academic outcomes of 12^{th} grade students



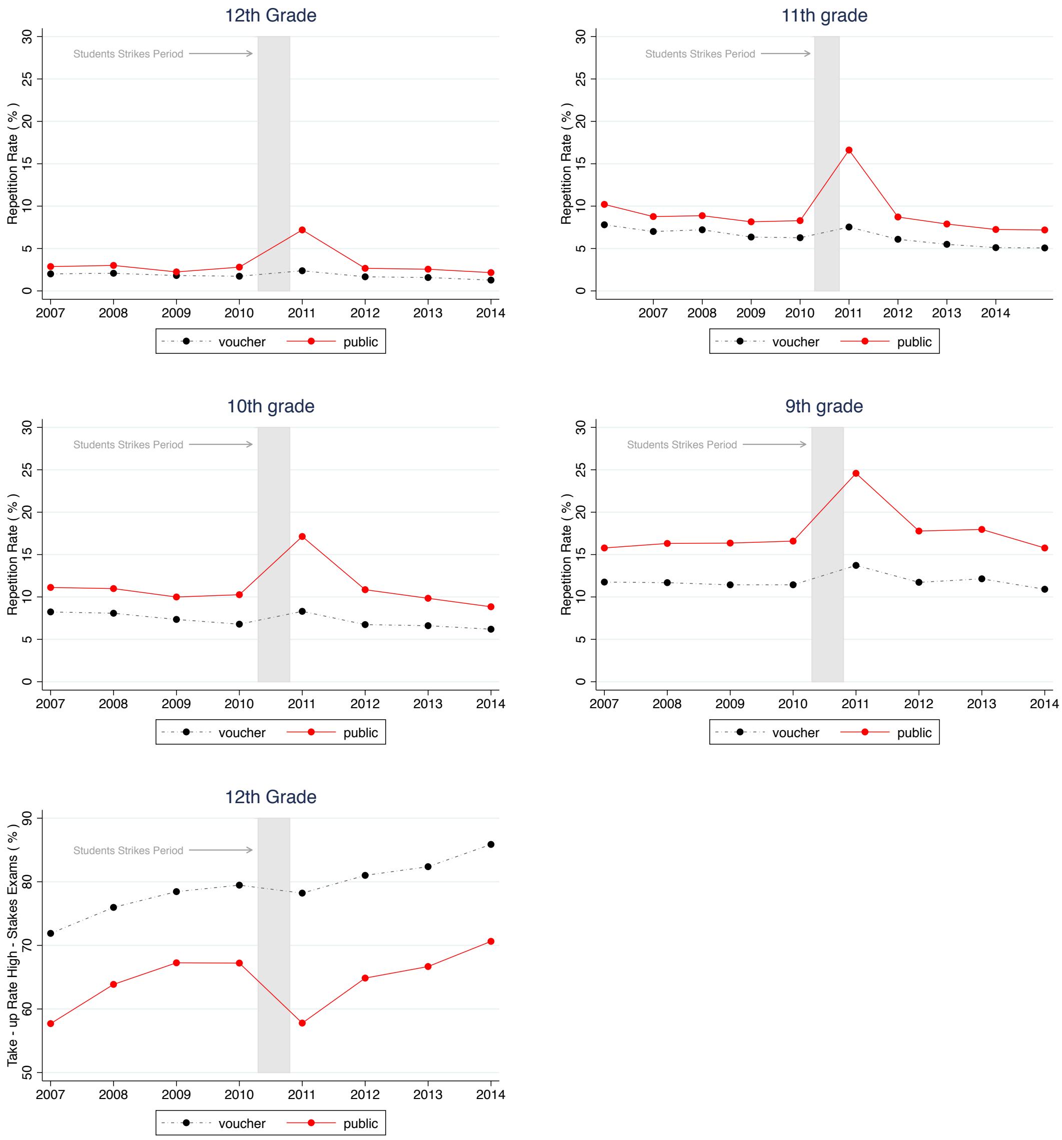
Notes: Top panel plots the average standardized score in the high-stakes math exam during the academic years 2007-2014. Middle panel plots the average standardized score in the high-stakes language exam during the academic years 2007-2014. Bottom panel plots the average enrollment rates in selective universities during the academic years 2007-2014. In order to facilitate the interpretation, levels in year 2007 are set to 0.

Figure A.7: Effect of the student strikes on 12^{th} grade public school students' academic outcomes (Intention to Treat (ITT))



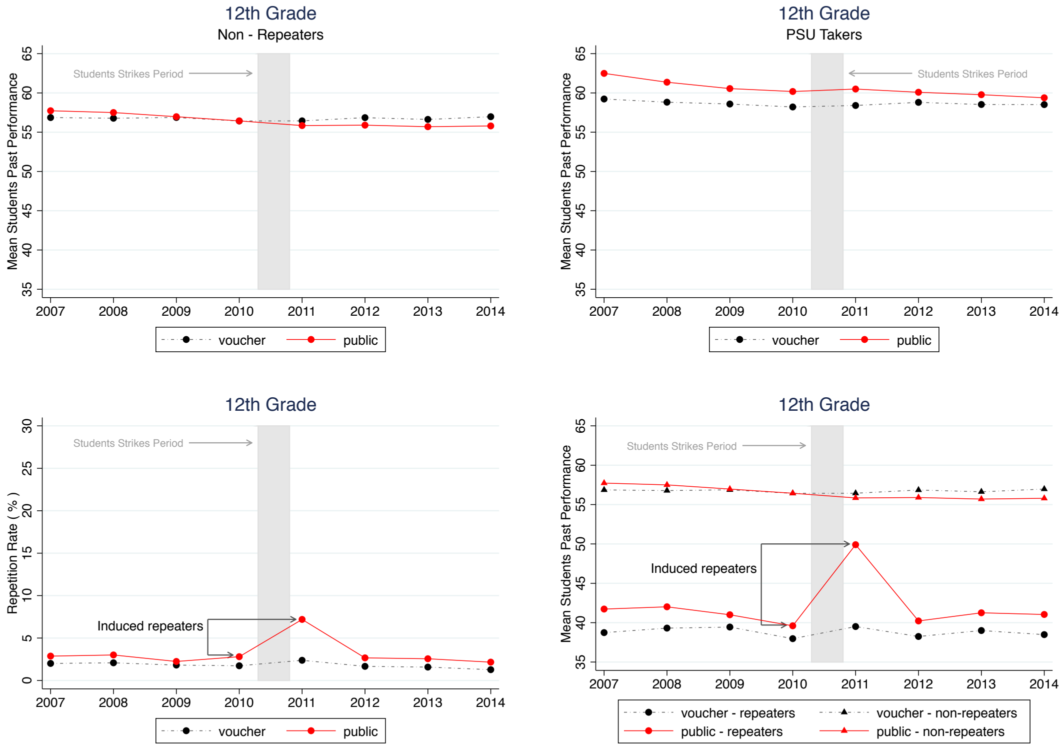
Notes: Top panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes math exam of 12^{th} grade students as a dependent variable. Middle panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes language exam of 12^{th} grade students as a dependent variable. Bottom panel plots parameter estimates of Equation (1.4), using the enrollment status in a selective university of 12^{th} grade students right after finishing secondary school as a dependent variable. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2014. I include school fixed effects and region $\times$ time fixed effects. At student level, I control for pre-strikes measures of individual students' past performance. I also include a gender dummy. In addition, I assign to students who switched schools in 2012, 2013 and 2014; their 2011 school.

Figure A.8: Repetition rates in grades 9^{th} to 12^{th} and high-stakes exams take-up rates



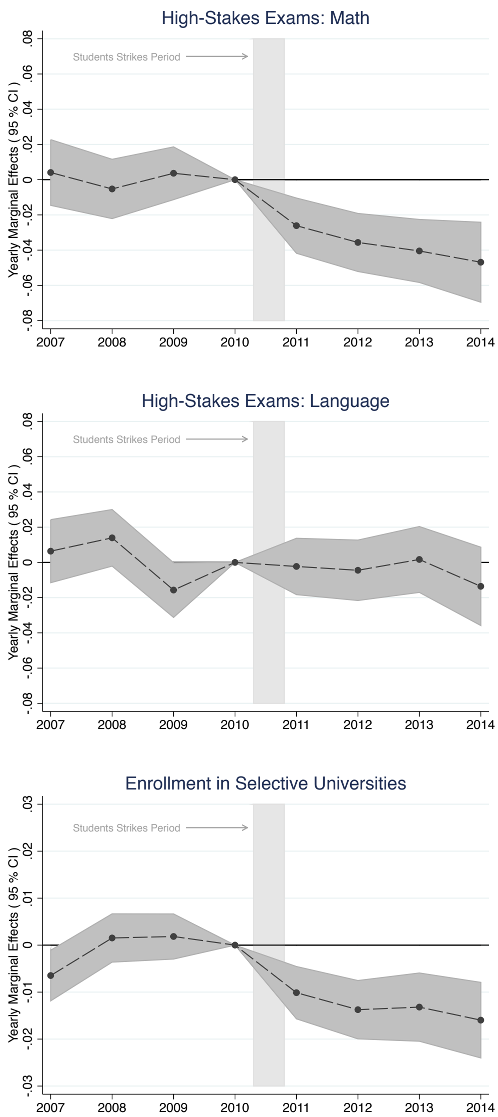
Notes: Top-left panel plots the average repetition rates in 12^{th} grade during the academic years 2007-2014. Top-right panel plots the average repetition rates in 11^{th} grade during the academic years 2007-2014. Middle-left panel shows the average repetition rates in 10^{th} grade during the academic years 2007-2014. Middle-right panel shows the average repetition rates in 9^{th} grade during the academic years 2007-2014. Bottom plots the average high-stakes exams' take-up rates during the academic years 2007-2014.

Figure A.9: Past performance of 12^{th} grade students and repetition rates



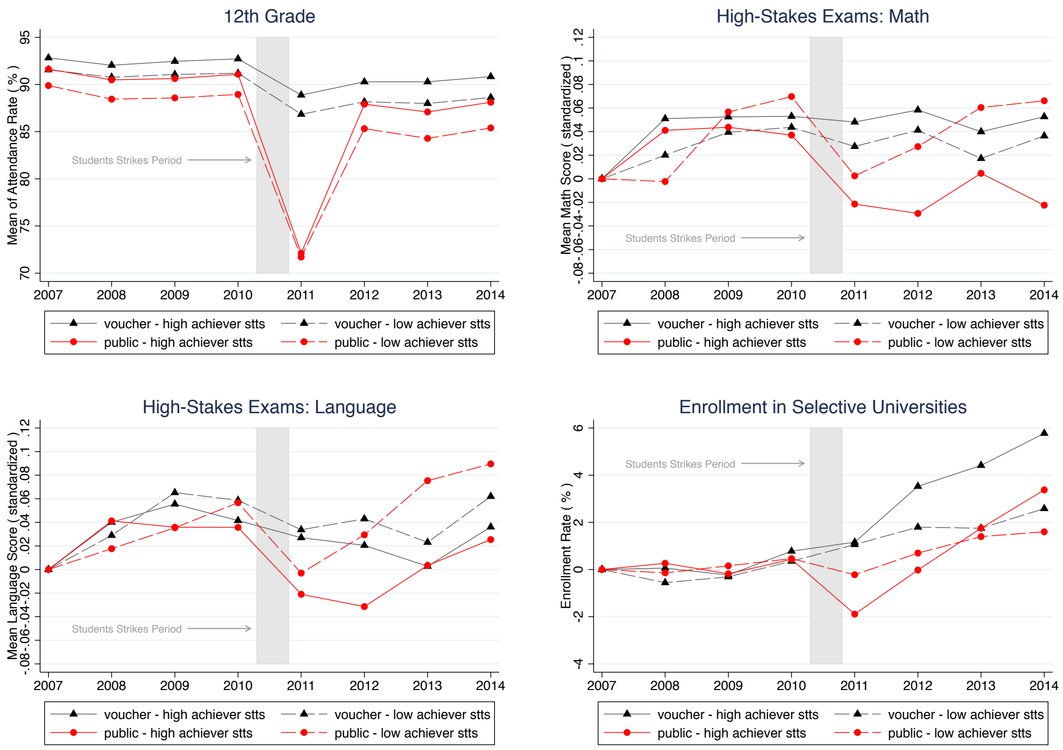
Notes: Top-left panel plots the average past performance of 12^{th} grade non-repeater students during the academic years 2007-2014, while the top-right panel plots the average past performance of 12^{th} grade non-repeater students that took the high-stakes exams in the same period. Bottom-left panel plots the average repetition rates in 12^{th} grade during the academic years 2007-2014. Bottom-right panel compares the past performance of 12^{th} grade students that repeated grade and those students that did not repeat grade during the academic years 2007-2014.

Figure A.10: Effect of the student strikes on 12^{th} grade public school students' academic outcomes (placing post-strikes repeaters back to their original cohorts)



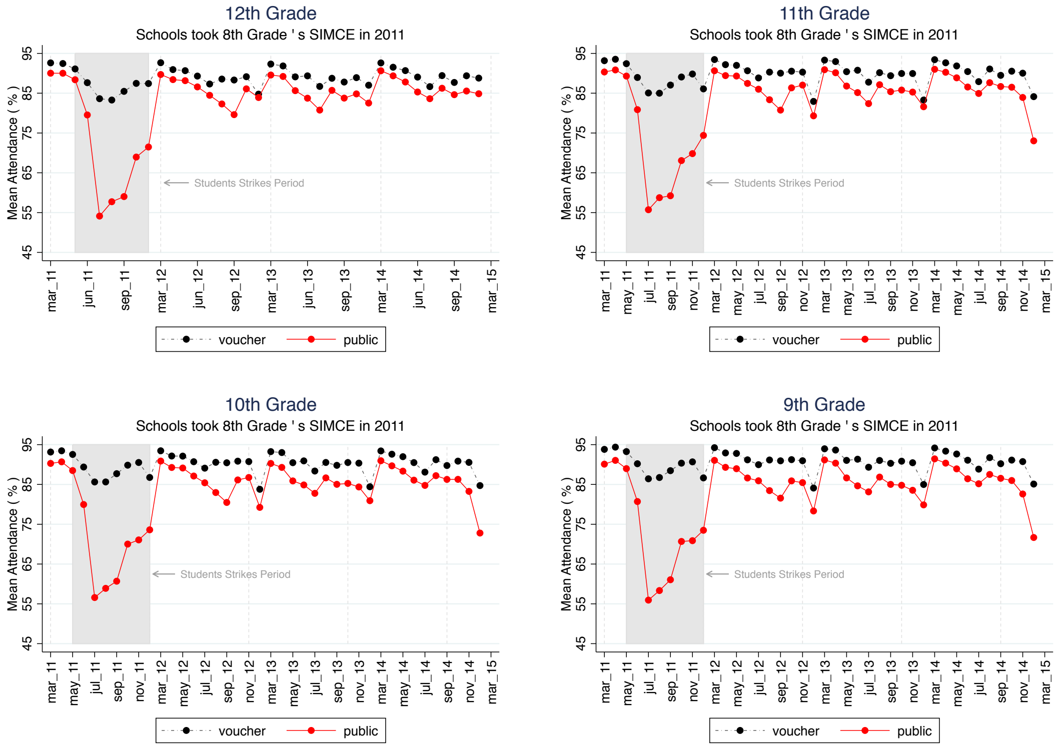
Notes: Top panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes math exam of 12^{th} grade students as a dependent variable. Middle panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes language exam of 12^{th} grade students as a dependent variable. Bottom panel plots parameter estimates of Equation (1.4), using the enrollment status in a selective university of 12^{th} grade students right after finishing secondary school as a dependent variable. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2014. I include school fixed effects and region $\times$ time fixed effects. At student level, I control for pre-strikes measures of individual students' past performance. I also include a gender dummy. In addition, I place post-strikes repeaters back to their original cohorts.

Figure A.11: Yearly attendance rates and academic outcomes of 12^{th} grade students by students' past performance



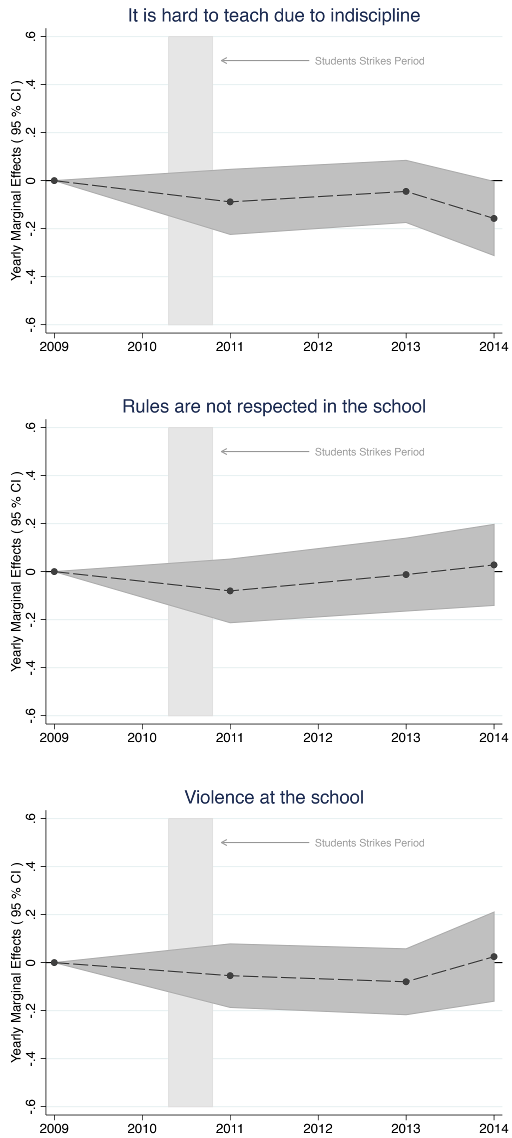
Notes: Top-left panel plots the average monthly attendance rates in 12^{th} grade during the academic years 2007-2014. Top-right panel plots the average standardized score in the high-stakes math exam during the academic years 2007-2014. Bottom-left panel shows the average standardized score in the high-stakes language exam during the academic years 2007-2014. Bottom-right panel shows the average enrollment rates in selective universities during the academic years 2007-2014. In order to facilitate the interpretation, levels in year 2007 are set to 0.

Figure A.12: Monthly attendance rates in grades 9^{th} to 12^{th} among the schools that offer both primary and secondary education and took the 8^{th} grade SIMCE test in 2011



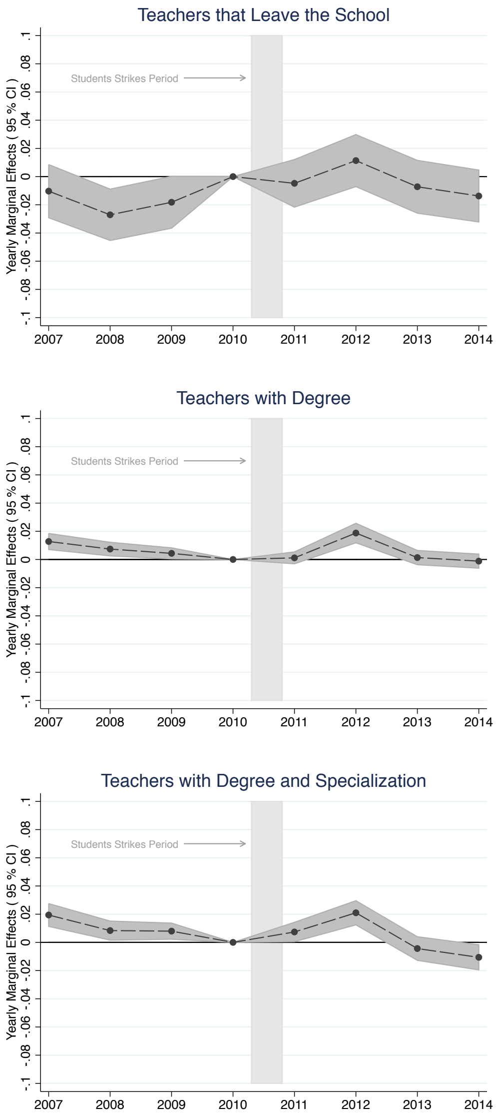
Notes: Top-left panel plots the average monthly attendance rates in 12^{th} grade during the academic years 2011-2014 among the schools that offer both primary and secondary education and took the 8^{th} grade SIMCE test in 2011. Top-right panel plots the average monthly attendance rates in 11^{th} grade during the academic years 2011-2014 among the schools that offer both primary and secondary education and took the 8^{th} grade SIMCE test in 2011. Bottom-left panel shows the average monthly attendance rates in 10^{th} grade during the academic years 2011-2014 among the schools that offer both primary and secondary education and took the 8^{th} grade SIMCE test in 2011. Bottom-right panel shows the average monthly attendance rates in 9^{th} grade during the academic years 2011-2014 among the schools that offer both primary and secondary education and took the 8^{th} grade SIMCE test in 2011.

Figure A.13: Effect of the student strikes on school environment



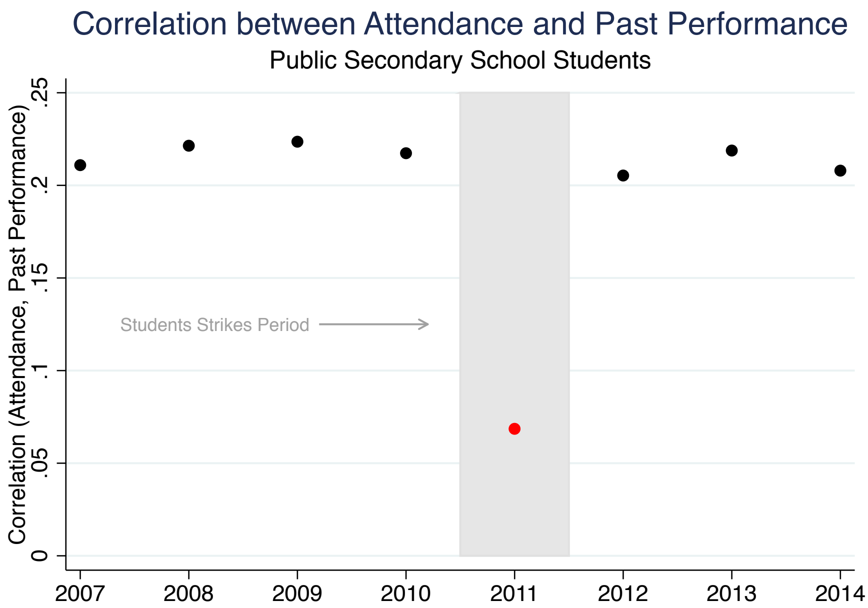
Notes: Top panel plots parameter estimates of Equation (1.4), using as a dependent variable the standardized score regarding the difficulty to teach in the school due to students' indiscipline behavior. Middle panel plots parameter estimates of Equation (1.4), using as a dependent variable the standardized score regarding the degree on which the rules in the school are respected by the students. Bottom panel plots parameter estimates of Equation (1.4), using as a dependent variable the standardized score regarding the degree of violence at the school. In the regressions, I use teacher-level data for years 2009, 2011, 2013 and 2014. I include school fixed effects and region $\times$ time fixed effects. At teacher level, I control for dummy variables regarding the subject taught by each teacher. I also include a gender dummy. All outcome variables are questions answered by the teachers regarding different situations at school level and these outcomes are standarized at year level.

Figure A.14: Effect of the student strikes on secondary school teachers



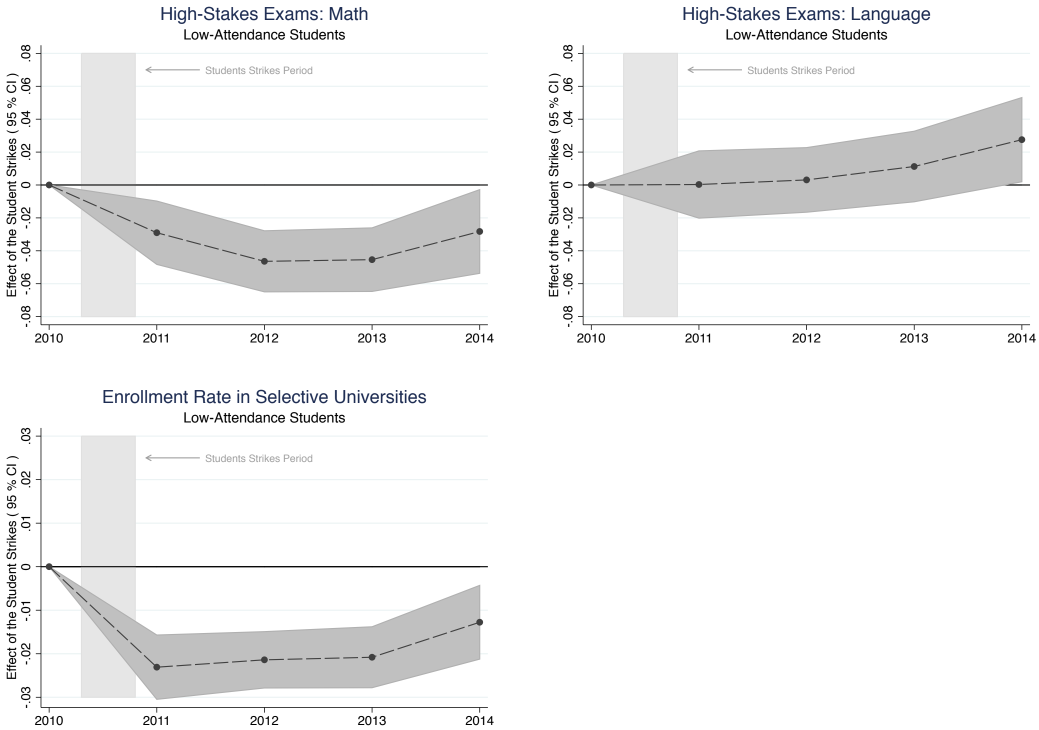
Notes: Top panel plots parameter estimates of Equation (1.4), using as a dependent variable an indicator variable that takes the value of 1 if the teacher leaves the school. Middle panel plots parameter estimates of Equation (1.4), using as a dependent variable an indicator variable that takes the value of 1 if the teacher holds a degree. Bottom panel plots parameter estimates of Equation (1.4), using as a dependent variable an indicator variable that takes the value of 1 if the teacher holds a degree and a specialization. In the regressions, I use yearly data at teacher level during the academic years 2007-2014. I include school fixed effects and region × time fixed effects. At techer level, I control for teachers' age and a gender dummy.

Figure A.15: Correlation between yearly school attendance and past performance of public secondary school students



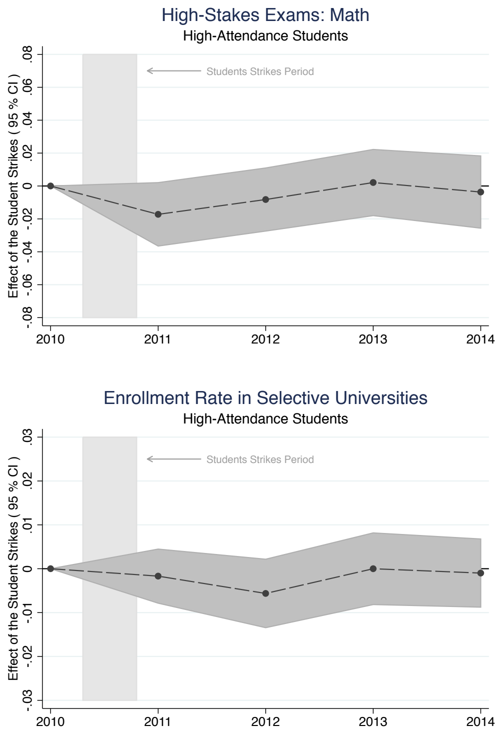
Notes: The figure plots the correlation between public secondary school students' past performance and their yearly attendance rate over the academic years 2007-2014.

Figure A.16: Effect of the student strikes on 12^{th} grade public school students' academic outcomes (public school students below median of the yearly attendance in 2011)



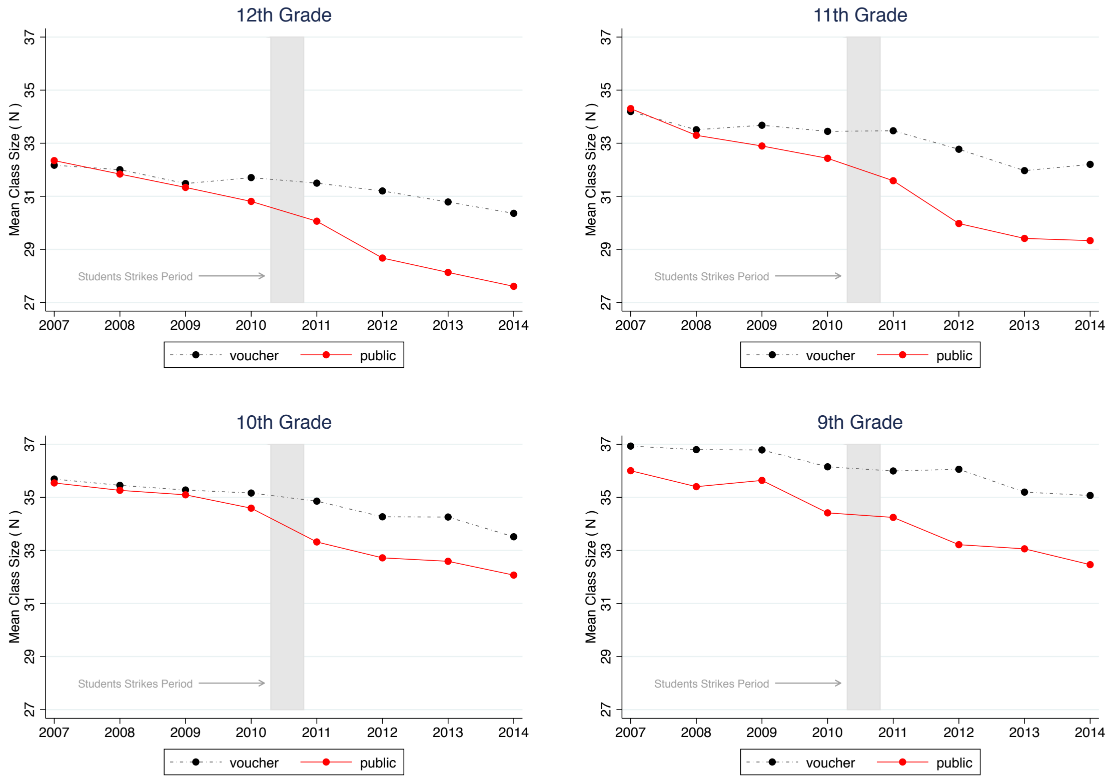
Notes: Top panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes math exam of 12^{th} grade students as a dependent variable. Middle panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes language exam of 12^{th} grade students as a dependent variable. Bottom panel plots parameter estimates of Equation (1.4), using the enrollment status in a selective university of 12^{th} grade students right after finishing secondary school as a dependent variable. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2014. For the academic year 2010, I keep in my sample all 12^{th} grade students in public and voucher schools. For the academic years 2011 onwards, I keep in my sample all 12^{th} grade students in voucher schools and 12^{th} grade students in public schools whose yearly attendance in 2011 was below the median of public school students' attendance distribution in their cohort that year. I include school fixed effects and region $\times$ time fixed effects. At student level, I control for pre-strikes measures of individual students' past performance. I also include a gender dummy. In addition, I place post-strikes repeaters back to their original cohorts.

Figure A.17: Effect of the student strikes on 12^{th} grade public school students' academic outcomes (public school students above median of the yearly attendance in 2011)



Notes: Top panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes math exam of 12^{th} grade students as a dependent variable. Middle panel plots parameter estimates of Equation (1.4), using the standardized score in the high-stakes language exam of 12^{th} grade students as a dependent variable. Bottom panel plots parameter estimates of Equation (1.4), using the enrollment status in a selective university of 12^{th} grade students right after finishing secondary school as a dependent variable. In the regressions, I use yearly repeated cross-section data at student level during the academic years 2007-2014. For the academic year 2010, I keep in my sample all 12^{th} grade students in public and voucher schools. For the academic years 2011 onwards, I keep in my sample all 12^{th} grade students in voucher schools and 12^{th} grade students in public schools whose yearly attendance in 2011 was above the median of public school students' attendance distribution in their cohort that year. I include school fixed effects and region $\times$ time fixed effects. At student level, I control for pre-strikes measures of individual students' past performance. I also include a gender dummy. In addition, I place post-strikes repeaters back to their original cohorts.

Figure A.18: Class size in grades 9^{th} to 12^{th}



Notes: Top-left panel plots the average class size in 12^{th} grade during the academic years 2007-2014. Top-right panel plots the average class size in 11^{th} grade during the academic years 2007-2014. Bottom-left panel shows the average class size in 10^{th} grade during the academic years 2007-2014. Bottom-right panel plots the average class size in 9^{th} grade during the academic years 2007-2014.

Tables

Table A.1: Yearly effect of the student strikes on students' academic outcomes

	(1) Math	(2) Language	(3) Enrollment
public × 2007	0.00381	0.00657	-0.00624**
	(0.00952)	(0.00913)	(0.00275)
public × 2008	-0.00550	0.0142*	0.00175
	(0.00860)	(0.00820)	(0.00264)
public × 2009	0.00342	-0.0155*	0.00211
	(0.00765)	(0.00794)	(0.00246)
public × 2011	-0.0333***	-0.0152*	-0.0144***
	(0.00809)	(0.00837)	(0.00297)
public × 2012	-0.0414***	-0.00744	-0.0156***
	(0.00831)	(0.00878)	(0.00319)
public × 2013	-0.0404***	-0.000298	-0.0127***
	(0.00897)	(0.00943)	(0.00352)
public × 2014	-0.0328***	0.00652	-0.00945**
	(0.0114)	(0.0112)	(0.00391)
Observations	1,033,409	1,033,409	1,429,263
R-squared	0.475	0.430	0.248
Student Level Controls	YES	YES	YES
School FE	YES	YES	YES
Region × Year FE	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are the standardized score in the high-stakes math exam, the standardized score in th high-stakes language exam and enrollment in a selective university, respectively. High-stakes exams scores are standardized within each year and enrollment is a dummy variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. Year dummy variables are intercated with *Public* and they take the value of 1 if the observation corresponds to that specific year, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects. *Public × 2010* is used as base category.

Table A.2: First-stage regressions for each grade of secondary school (whole cohort of students)

	(1) Abs_9^{th}	(2) Abs_10^{th}	(3) Abs_11^{th}	(4) Abs_12^{th}
public × 2011	12.46***	12.55***	13.27***	14.85***
	(0.789)	(0.770)	(0.849)	(0.903)
Observations	908,511	910,836	914,123	922,582
R-squared	0.356	0.363	0.390	0.399
Student Level Controls	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES
F-test instrument	249.6	266	244.1	270.1

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level of the whole cohort of students during the academic years 2007-2010. For estimations regarding 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the whole cohort of 12^{th} grade students corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. Outcome variables are school absenteesim in 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects.

Table A.3: First-stage regressions for each grade of secondary school (only students that sat the high-stakes exams)

Outcome: Yearly attendance rate	(1) 9^{th} Grade	(2) 10^{th} Grade	(3) 11^{th} Grade	(4) 12^{th} Grade
public × 2011	14.08***	14.03***	14.88***	14.91***
	(0.949)	(0.936)	(1.029)	(1.098)
Observations	654,737	650,149	649,956	646,906
R-squared	0.372	0.372	0.401	0.401
Student Level Controls	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES
F-test instrument	220.3	224.5	209	184.3

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level of the students that sat the high-stakes exams during the academic years 2007-2010. For estimations regarding 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the cohort of 12^{th} grade students that sat the high-stakes exams corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. Outcome variables are school absenteesim in 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects.

Table A.4: Effect of school absenteeism in each secondary school's grade on math

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
absenteeism in 9^{th} (%)	-0.00208***	-0.00325***						
	(0.000211)	(0.000702)						
absenteeism in 10^{th} (%)			-0.00326***	-0.00277***				
			(0.000197)	(0.000610)				
absenteeism in 11^{th} (%)					-0.00383***	-0.00284***		
					(0.000204)	(0.000551)		
absenteeism in 12^{th} (%)							-0.00414***	-0.00234***
							(0.000190)	(0.000496)
Observations	654,737	654,737	650,149	650,149	649,956	649,956	646,906	646,906
R-squared	0.474	0.474	0.484	0.484	0.483	0.483	0.486	0.485
Student Level Controls	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES
Region $\times$ Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2010. For estimations regarding 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the cohort of 12^{th} grade students corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. Outcome variable is the standardized score in the high-stakes math exam. High-stakes math exam's score is standardized within each year. Attendance rate on each grade is measured in percentage points and goes from 0 to 100. The instrument for school absenteeism is ($Public_s \times 2011_t$), which is a dummy variable that takes the value of 1 if the student was attended a public school in 2011, 0 otherwise. Odd columns report the OLS estimates for 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. Even columns report their counterpart IV estimates. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time $\times$ region fixed effects.

Table A.5: Effect of school absenteeism in each secondary school's grade on language

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
absenteeism in 9^{th} (%)	0.000187	-0.00121*						
	(0.000212)	(0.000662)						
absenteeism in 10^{th} (%)			-0.000792***	-0.000213				
			(0.000199)	(0.000642)				
absenteeism in 11^{th} (%)					-0.00119***	-0.000684		
					(0.000172)	(0.000574)		
absenteeism in 12^{th} (%)							-0.00159***	-0.00121**
							(0.000165)	(0.000513)
Observations	654,737	654,737	650,149	650,149	649,956	649,956	646,906	646,906
R-squared	0.435	0.435	0.436	0.436	0.441	0.441	0.443	0.443
Student Level Controls	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2010. For estimations regarding 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the cohort of 12^{th} grade students corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. Outcome variable is the standardized score in the high-stakes language exam. High-stakes language exam's score is standardized within each year. Attendance rate on each grade is measured in percentage points and goes from 0 to 100. The instrument for school absenteeism is ($Public_s \times 2011_t$), which is a dummy variable that takes the value of 1 if the student was attended a public school in 2011, 0 otherwise. Odd columns report the OLS estimates for 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. Even columns report their counterpart IV estimates. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects.

Table A.6: Effect of school absenteeism in each secondary school's grade on university enrollment

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
absenteeism in 9^{th} (%)	-0.000809***	-0.000845***						
	(0.000070)	(0.000302)						
absenteeism in 10^{th} (%)			-0.00137***	-0.000964***				
			(0.000082)	(0.000262)				
absenteeism in 11^{th} (%)					-0.00151***	-0.00124***		
					(0.000085)	(0.000242)		
absenteeism in 12^{th} (%)							-0.00169***	-0.00106***
							(0.000078)	(0.000168)
Observations	908,511	908,511	910,836	910,836	914,123	914,123	922,582	922,582
R-squared	0.250	0.250	0.251	0.251	0.254	0.254	0.253	0.253
Student Level Controls	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2010. For estimations regarding 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, I also include the cohort of 12^{th} grade students corresponding to the academic year 2014, 2013, 2012 and 2011, respectively. The previous procedure is discussed more in detail in Section 1.7.2. Outcome variable is enrollment in a selective university, which is an indicator variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. Attendance rate on each grade is measured in percentage points and goes from 0 to 100. The instrument for school absenteeism is ($Public_s \times 2011_t$), which is a dummy variable that takes the value of 1 if the student was attended a public school in 2011, 0 otherwise. Odd columns report the OLS estimates for 9^{th} grade, 10^{th} grade, 11^{th} grade and 12^{th} grade, respectively. Even columns report their counterpart IV estimates. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects.

Table A.7: Effect of secondary school absenteeism on academic outcomes (controlling for class size)

		Math		Language			Enrollment	
	(1) Absenteeism	(2) OLS	(3) IV	(4) OLS	(5) IV	(6) Absenteeism	(7) OLS	(8) IV
public × 2011	4.184***					3.998***		
	(0.266)					(0.231)		
absenteeism in secondary (%)		-0.00873***	-0.00835***	-0.00136***	-0.00170		-0.00367***	-0.00293***
		(0.000337)	(0.00160)	(0.000283)	(0.00161)		(0.000141)	(0.000583)
Observations	1,033,404	1,033,404	1,033,404	1,033,404	1,033,404	1,429,263	1,429,263	1,429,263
R-squared	0.390	0.478	0.478	0.432	0.432	0.386	0.250	0.250
Student Level Controls	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	YES	YES	YES	YES	YES
F-test on instrument	248.1					298.7		

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. All standard errors are clustered at school level and reported in parenthesis. This table only shows estimates of interest. The data contains repeated cross-section information at students level during the academic years 2007-2014. Outcome variables are school absenteesim during secondary school education, the standardized score in the high-stakes math exam, the standardized score in the high-stakes language exam and enrollment in a selective university, respectively. Column 1 reports the first-stage estimates only considering the population of students that sat the high-stakes exams, while column 6 reports the first-stage estimates considering the whole population of students. Columns 2, 4, and 7 report the OLS estimates regarding the effect of school absenteesim during secondary school education on academic outcomes, while columns 3, 5, and 8 reports the IV estimates. School absenteesim during secondary school education is measured in percantage points, and therefore it is a continuous variable that can take values from 0 to 100. High-stakes exams scores are standardized within each year. Enrollment is a dummy variable that takes the value of 1 if the student was enrolled in a selective university right after finishing 12^{th} grade. *Public* is a dummy variable that takes the value of 1 if the student attends to a public school, 0 otherwise. *2011* is a dummy variable that takes the value of 1 if the observation corresponds to year 2011, 0 otherwise. Al class-level, I control for class size. Student-level controls include a measure of students' ability unaffected by the strikes (the rank position of the students within their class 4 years before 12^{th} grade, which is a continuous variable that goes between 0 and 100) and a gender dummy. All regressions include school fixed effects and time × region fixed effects. In addition, for each student in cohorts after 2011, I assign them the school (and therefore, also the region) she attended in 2011.

Characterizing students that were *induced* to repeat by the student strikes

Let me define:

A_i as the measure of student i past performance.

Y_i as a dummy variable that takes the value of 1 if student i is observed in year 2011.

I_i as a dummy variable that takes the value of 1 if student i was *induced* to repeat by the student strikes.

R_i as a dummy variable that takes the value of 1 if student i repeated the grade.

S_i as a dummy variable that takes the value of 1 if student i attended a public school.

By conditional expectations:

$$\begin{aligned}
\underbrace{\mathbb{E}\left[A_i \,|\, R_i = 1,\ Y_i = 1,\ S_i = 1\right]}_{\substack{\text{mean students' past performance}\\ \text{of repeaters in 2011}}} = \; & \underbrace{\mathbb{E}\left[A_i \,|\, I_i = 0,\ R_i = 1,\ Y_i = 1,\ S_i = 1\right]}_{\substack{\text{mean students' past performance}\\ \text{of \textit{always} repeaters in 2011}}} \times \\
& \underbrace{Pr(I_i = 0 \,|\, R_i = 1,\ Y_i = 1,\ S_i = 1)}_{\substack{\text{proportion of \textit{always}}\\ \text{repeaters in 2011}}} + \\
& \underbrace{\mathbb{E}\left[A_i \,|\, I_i = 1,\ R_i = 1,\ Y_i = 1,\ S_i = 1\right]}_{\substack{\text{mean students' past performance}\\ \text{of \textit{induced} repeaters in 2011}}} \times \\
& \underbrace{Pr(I_i = 1 \,|\, R_i = 1,\ Y_i = 1,\ S_i = 1)}_{\substack{\text{proportion of \textit{induced}}\\ \text{repeaters in 2011}}}
\end{aligned} \tag{A.1}$$

Where,

$$Pr(I_i = 1 \,|\, R_i = 1,\ Y_i = 1,\ S_i = 1) = 1 - Pr(I_i = 0 \,|\, R_i = 1,\ Y_i = 1,\ S_i = 1) \tag{A.2}$$

By conditional probability:

$$Pr(I_i = 0 \,|\, R_i = 1,\, Y_i = 1,\, S_i = 1) \;=\; \frac{Pr(I_i = 0 \,\cap\, R_i = 1 \,|\, Y_i = 1,\, S_i = 1)}{Pr(R_i = 1 \,|\, Y_i = 1,\, S_i = 1)} \qquad \text{(A.3)}$$

Consider the following assumptions:

A1. *Timing of the strikes*
Strikes only affected the behaviour of public school students in 2011:
$Pr(I_i = 1 \,|\, Y_i = 0,\, S_i = 1) = \mathbb{E}\,[I_i \,|\, Y_i = 0,\, S_i = 1] = 0$

A2. *Proportion of repeaters*
The proportion of always *repeaters in public schools is stable over time:*

$Pr(I_i = 0 \,\cap\, R_i = 1 \,|\, Y_i = 1,\, S_i = 1) \approx$
$Pr(R_i = 1 \,|\, Y_i = 0,\, S_i = 1) = \mathbb{E}\,[R_i \,|\, Y_i = 0,\, S_i = 1]$

A3. *Past performance of repeaters*
The mean past performance of always *repeaters in public schools is stable over time:*
$\mathbb{E}\,[A_i \,|\, I_i = 0,\, R_i = 1,\, Y_i = 1,\, S_i = 1] \approx \mathbb{E}\,[A_i \,|\, R_i = 1,\, Y_i = 0,\, S_i = 1]$

Assumption 1 comes from the fact that the student strikes only took place during 2011. Assumptions 2 and 3 are based on the trends presented in the top-left and the bottom-left panel of figure A.9.

From the data,

$\mathbb{E}\,[A_i \,|\, R_i = 1,\, Y_i = 0,\, S_i = 1] = 41.01$

$\mathbb{E}\,[R_i \,|\, Y_i = 0,\, S_i = 1] = 0.0263$

$Pr(R_i = 1 \,|\, Y_i = 1,\, S_i = 1) = \mathbb{E}\,[R_i \,|\, Y_i = 1,\, S_i = 1] = 0.0719$

$\mathbb{E}\,[A_i \,|\, R_i = 1,\, Y_i = 1,\, S_i = 1] = 49.9$

$\mathbb{E}\,[A_i \,|\, R_i = 0,\, Y_i = 1,\, S_i = 1] = 58.84$

Combining Equations (A.1), (A.2) and (A.3), using assumptions 1, 2 and 3, and the data; implies that for students in public schools:

$$\underbrace{41.01}_{\substack{\text{mean students'}\\ \text{past performance}\\ \text{of } \textit{always}\\ \text{repeaters in 2011}}} \times \underbrace{\frac{0.0263}{0.0719}}_{\substack{\text{proportion}\\ \text{of } \textit{always}\\ \text{repeaters}\\ \text{in 2011}}} + \underbrace{y}_{\substack{\text{mean students'}\\ \text{past performance}\\ \text{of } \textit{induced}\\ \text{repeaters in 2011}}} \times \underbrace{\frac{(0.0719 - 0.0263)}{0.0719}}_{\substack{\text{proportion}\\ \text{of } \textit{induced}\\ \text{repeaters}\\ \text{in 2011}}} \approx \underbrace{49.9}_{\substack{\text{mean students'}\\ \text{past performance}\\ \text{of repeaters}\\ \text{in 2011}}}$$
$$\implies y \approx 55.03 \approx \underbrace{55.84}_{\substack{\text{mean students'}\\ \text{past performance}\\ \text{of non-repeaters}\\ \text{in 2011}}} \tag{A.4}$$

Chapter 2

Estimating Value-Added Models for Doctoral Teaching Assistants: Evidence from a Random Assignment Procedure at a UK University

2.1 Introduction

The value-added (VA) approach is the most prominent method for evaluating teacher quality. This is based on measuring teacher impacts on students' test scores. Research using VA models indicates that teacher quality is an important determinant of student achievement in primary and secondary education (Rockoff (2004), Rivkin et al. (2005), Aaronson et al. (2007), Kane and Staiger (2008), Chetty et al. (2014b)). Evidence for how the quality of instruction affects students' academic outcomes is rarer at the post-secondary level. This is because exams are not standardized, and students generally choose their course work and their teachers, leading to an endogenous match that can impose a significant bias in teachers' value-added estimates (Rothstein (2010b)). Moreover – even under random assignment – VA estimates can be eroded if teachers "teach to the test" or favor their students by simplifying the assessments or the content of the course, or by inflating academic tests scores

(Carrell and West (2010), Chetty et al. (2014b)).

This chapter focuses on the Economics Department of a UK university where, for each core module, undergraduate (UG) students are randomly allocated to seminar classes taught by different doctoral teaching assistants (DTAs). Doctoral teaching assistants are widely-employed in the post-secondary sector and they perform various duties that can impact UG students' success in the course (Lusher et al. (2015)): they host weekly discussion classes, solving questions, exercises, and problem sets. DTA-student relationships are more like a peer-to-peer interaction, since the age gap between undergraduates and DTAs is usually small.[1] Furthermore, with student-professor ratios increasing, DTA's influence on students' academic achievement is likely to grow over time (Cuseo (2007), Kokkelenberg et al. (2008), Schanzenbach (2014)). Beside their intrinsic policy relevance, we argue that DTAs are an informative group to study. DTAs do not decide what is taught, write exams or systematically mark their students' scripts. Arguably, these features allow estimation of value-added models that are less likely to confound "teaching quality" with unrelated factors (Carrell and West (2010), Chetty et al. (2014b)).

Our dataset includes UG students' demographics, courses and classes attended, timing of the class, and identifiers of the DTA assigned to each class. An exclusive feature of the data is to provide marks for separate types of evaluation of the course taught by the DTA. In particular, we have marks for: (i) assessments carried out during the term, which have a low weight in the total mark; (ii) tests taken at the end of each term that carry a slightly higher weight; (iii) end of year exams, which account for about 80% of the total mark. There is usually a non-teaching break before the final exams that allows students to revise the material covered during the entire course.

The empirical approach is validated by tests confirming the reliability of the random allocation of UG students to classes. The value-added estimates show that a one standard deviation change in DTA quality increases students' scores by around 8.5 percent of a standard deviation (σ). This result is robust to the inclusion of type

[1] Several studies have focused on the potential benefits of peer-based mentoring and tutoring. For example, Castleman and Page (2015) find that near-aged peer mentors in college who sent text messages during the summer to college-intending high school graduates substantially increased college enrollment.

of evaluation fixed effects, students' characteristics, class characteristics, peer characteristics, and a set of fixed effects for the date/time of the class. The magnitude of the estimates is similar to other VA studies in post-secondary education exploiting random assignment of students to classes (Carrell and West (2010), Braga et al. (2016)).

The granularity of the data allows for a novel estimation of the within-course VA dynamics. This reveals a startling decay: Estimates are larger for assessments taken during the course (16% σ), drop for end-of-term tests (6% σ), and are not statistically different from zero for final exams, implying that there is no variation in DTA effectiveness for the final exams. These results are thoroughly examined: robustness checks confirm that the lack of persistence of the VA estimates is genuine, and it is not driven by changes in the sample composition, nor by outliers.

While previous research in post-secondary education has detected long-term teacher-effects for subsequent courses (taught by a different teacher) and even in the labor market (Braga et al. (2016)), in this chapter we find a sharp VA decay through the academic year for separate evaluations of the same course taught by the same teacher. In general, the (lack of) persistence of the teacher effects could reflect two main mechanisms (Jacob et al. (2010)): (i) Students may engage in potentially endogenous subsequent investments, which either mitigate or exacerbate the consequences of the allocation to a particular teacher; (ii) students may forget information or lose skills that they acquired from a particular teacher. The underlying mechanisms suggest that the decay might be related to students' endogenous investments. For example, students could compensate for being assigned to a low-value-added DTA by working harder in the non-directed study period before the final exam. This increase in effort might be amplified by the high weight of the final exam in the course final mark. Nevertheless, we cannot rule out that the lack of variation in DTAs effectiveness detected in final exams is consistent with students rapidly losing skills that they acquired when interacting with a DTA.

This chapter contributes to the literature estimating VA models in post-secondary education. Two studies have exploited the random assignment of students to teachers. Carrell and West (2010) examine students at the U.S. Air Force Academy. They find that a one standard deviation change in professor quality increases students' scores by around 0.05σ. Instructors that were better at improving contemporary performance received higher teacher evaluations but were less successful at promoting "deep-learning", as indicated by student performance in subsequent courses. Braga et al. (2016) estimate teacher effects at Bocconi University. They find significant variation in teacher effectiveness, roughly 0.04σ both for academic and labor market outcomes. The professors who were best at improving the academic achievement of their best students were also the ones who boosted their earnings the most. Focusing on large, introductory courses at a Canadian research university, Hoffmann and Oreopoulos (2009) find the standard deviation of professor effectiveness measured by course marks is no larger than 0.08σ. Other recent studies have concluded that instructors play a larger role in student success. Bettinger et al. (2014) examine instructor effectiveness using data from DeVry University, a large for-profit institution in which the average student takes two-thirds of her courses online. They find that being taught by an instructor that is 1σ more effective improves student course marks by about 0.18σ to 0.24σ. Among instructors of economics, statistics and computer science at an elite French public university, Brodaty and Gurgand (2016) find that a 1σ increase in teacher quality is associated with a 0.14σ to 0.25σ increase in student test results. De Vlieger et al. (2017) focus on the University of Phoenix, a large for-profit university that offers both online and in person courses in a wide array of fields and degree programs. A 1σ increase in instructor quality is associated with 0.3σ increase in marks in current course and a 0.2σ increase in marks in the subsequent course in the math sequence.

Our study also adds on the literature on teaching assistants in post-secondary education. Few related papers study how the origin and ethnicity of graduate teaching assistants (TAs) affect student performance. Lusher et al. (2015) examine the role of graduate TAs' ethnicity and find that students' marks increase when they are assigned to same-ethnicity graduate TAs. Borjas (2000) finds that foreign-born TAs negatively affect student marks. Fleisher et al. (2002) find that foreign-born graduate TAs have negligible effects on student marks which, in some circumstances, can even be positive. Bettinger et al. (2016) look at the effect of having a PhD student as a full instructor on students' subsequent major choices. They find that students are more likely to major in a subject if the first courses in that subject are taught by a PhD student. Feld et al. (2017) study the effect of being assigned to a bachelor or a masters student as an instructor. Student instructors are almost as effective as senior instructors at improving academic achievement and labor market outcomes.

This chapter unfolds as follows. Section 2.2 describes the institutional background and the data. Section 2.3 introduces the empirical design and presents four tests on the random assignment procedure of students to classes performed by the University administration. Section 2.4 presents the results, robustness checks, and heterogeneity analysis. Section 2.5 concludes.

2.2 Institutional Background

The analysis focuses on undergraduate students enrolled in a highly ranked Economics Department of a UK University.[2] Students' acceptance to the program is based on a minimum of A*AA in the A-level examination, with a compulsory mathematics component. Admission is highly competitive: only 9% of the students that sat A-level exams in England during the academic year 2011/2012 met this threshold.[3] The undergraduate program is 3 years with a mix of core and optional courses. Core courses might differ across the 5 main bands of the program.[4]Each academic year has three ten-week terms. Students generally take five courses. In the first two terms there are one or two lectures per week, taught by faculty members.[5]

PhD students in Economics are typically employed as teaching assistants.[6] They

[2]This Department is regularly among the top five UK Economics Departments in different rankings, including QS World University Ranking, Shangai University Ranking, IDEAS/RePEc Ranking, Times Higher Education Ranking, among others.

[3]International students who studied outside the UK need to score 38 points in the International Baccalaureate, including 6 points in Higher Level Mathematics. This appears to be an equally selective threshold. More information on the international baccalaureate can be found here.

[4]The 5 main bands of the undergraduate program are Economics; Economics and Industrial Organization; Economics and Economic History; Philosophy, Politics and Economics and Mathematics and Economics.

[5] For a small minority of courses, there are also lectures in the third term.

[6] Selection for PhD students in the Department of Economics involves screening by an ad-hoc committee. Prospective doctoral students submit a personal statement, academic transcripts, English certificate, GRE and two academic references. While there is no published threshold for doctoral candidates, on average each year the Department receives 230 applications and accepts 15-20 students. This puts the acceptance rate below 9%.

teach weekly seminars for small groups of undergraduate students. Attendance is mandatory and the seminars go for one hour. Within each course, DTAs follow the same syllabus and use identical academic material including questions, exercises, and problem sets. DTAs do not choose the academic material, nor do they prepare assessments, tests, or final exams. The course leader, who is faculty, designs the course, prepares teaching material and writes assessments. PhD students do mark scripts. However, two features reduce the concerns that DTAs might systematically inflate their students' marks. First, assessments and end-of-term tests are randomly allocated among all the DTAs teaching the course and final exams are randomly allocated among PhD students in the Department of Economics, who are not necessarily DTAs. This means that DTAs do not systematically mark their students' scripts. Second, university regulation requires that all official scripts (assessments, end-of-term tests, final exams) be anonymous. The sole mode of recognition is a 7-digit identifier.

2.2.1 Data

The sample includes data from the 2008/09 to the 2011/12 academic year. We focus on 8 core courses: Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy (1^{st} year courses); Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (2^{nd} year courses). These courses allow an unbiased estimation of the DTAs value-added measures because students in these courses are randomly allocated to seminars, and they employ multiple doctoral students as teaching assistants. The randomness of the allocation procedure will be tested in Section 2.3.1.

Table 2.1 provides summary statistics for the sample. There are 2,189 students: 40.1% are female; 44% come from overseas; 81.7% are enrolled in the Economics Department. The sample includes 66 DTAs' identifiers. For confidentiality reasons, no other information is available for doctoral teaching assistants. On average, each course employs 4 DTAs, has 2 assessments, 1 test and 1 final exam.[7] We have 24,914 evaluations measuring students' academic achievement against a set of course-

[7]All courses have a final exam. Macroeconomics I and Microeconomics II do not have mid-term assessments. The following 4 courses have end-of-term tests: Macroeconomics I, Microeconomics I, Microeconomics II, and Econometrics.

specific predetermined learning outcomes: 11,859 assessments; 6,542 end-of-term tests and 6,513 final exams. The mark scale goes from 0 to 100.[8] The data also provide information on the weight of each evaluation in the final mark of the course. The sample includes 589 classes, with an average size of nearly 14 students. Close to half of the classes are taught in the morning; 35.3% are on Mondays, while classes on Wednesdays account for 10.5% of the seminars. Finally, we have information on average students' evaluations of the DTA for each course taught. On a scale of 1 to 5, the average DTAs's score is 4 (where 1 is poor and 5 is excellent).

2.3 Empirical Design

The empirical strategy is set to examine the variation in student academic performance across DTAs teaching different classes in the same course, in a given year. The value-added model is defined as follows:

$$y_{eskdcy} = \gamma_e + \lambda_{cy} + x'_{kdcy} \times \beta_1 + z'_{skdcy} \times \beta_2 + \alpha_d + \epsilon_{eskdcy} \tag{2.1}$$

The outcome variable y_{aikjct} is the mark in the evaluation e obtained by the student s of class k in course c taught by the DTA d during the academic year y. We standardize the mark (within type of evaluation/course/year) to have a mean of zero and a standard deviation of one. Course $\times$ Year fixed effects (λ_{cy}) account for course/year specific shocks that are common across classes but independent across DTAs. The vector z_{skdcy} corresponds to a set of student-level controls, including a gender dummy, a dummy variable regarding student's overseas status, and a dummy variable indicating the enrollment in the Department of Economics. Class-level controls are contained in the vector x_{kdcy}. This includes class size, a dummy variable that takes the value of 1 if the class was taught in the morning, and a set of dummy variables for the day of the week on which the class was taught. This vector also includes the share of peers in the class that are: (i) Female, (ii) with overseas status, and (iii) enrolled in the Department of Economics. Evaluations fixed effects

[8]Figure B.1 in the appendix shows the distribution of marks by type of evaluation, also reporting mean and standard deviation.

(γ_e) control for time-invariant unobserved heterogeneity across different types of evaluations. Robust standard errors are clustered at the DTA level; ϵ_{eskdcy} is the student-specific stochastic error term.

Table 2.1: Descriptive Statistics

	Total	Mean	Standard Deviation
Course level:			
Number of courses	8		
Number of courses / year	32		
Number of DTAs		4.06	1.76
Number of classes		18.41	3.02
Number of students		207.78	60.96
Number of assessments		2.06	1.83
Number of midterms		0.91	1.00
Number of exams		1.00	0.00
Doctoral Teaching Assistant (DTA) level:			
Number of DTAs	66		
Number of courses		1.97	1.14
Number of classes		8.92	6.31
Number of assessments		179.68	154.33
Number of midterms		99.12	113.69
Number of exams		98.68	70.77
Students evaluation		4.00	0.49
Class level:			
Number of classes	589		
Class size		13.91	4.81
Morning (%)		53.48	
Monday (%)		35.31	
Tuesday (%)		15.96	
Wednesday (%)		10.53	
Thursday (%)		21.56	
Friday (%)		16.64	
Student level:	2,189		
Female (%)		40.11	
Overseas (%)		44.04	
Enrolled in econ. Dept. (%)		81.73	
Evaluation level:			
Number of assessments	11,859		
Number of tests	6,542		
Number of exams	6,513		
Weight of assessments (%)		6.05	4.21
Weight of tests (%)		8.64	1.64
Weight of exams (%)		80.23	12.84

Notes: The sample includes the following courses: Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2 and, Econometrics (from 2008/09 to 2011/12).

The term α_d represents the unobserved heterogeneity among the DTAs. This quantifies the individual contribution of each DTA to students' academic achievements. The within-course variance of this individual contribution is the main parameter of interest, and we call this the value-added estimate. Following Rockoff (2004), Carrell and West (2010), and De Vlieger et al. (2017), Equation (2.1) is estimated as a random intercept model. The following assumptions are imposed:

1. α_d are independent and identically distributed: $\alpha_d \sim \mathcal{N}(0,\, \sigma^2_\alpha)$;
2. ϵ_{eskdcy} are independent and identically distributed: $\epsilon_{eskdcy} \sim \mathcal{N}(0,\, \sigma^2_\epsilon)$;
3. α_d and ϵ_{eskdcy} are independent of each other.

The deterministic component of the model ($\gamma_e + \lambda_{cy} + x'_{kdcy} \times \beta_1 + z'_{skdcy} \times \beta_2$) is estimated by feasible Generalized Least Squared (GLS), considering the structure placed in the variance of both α_d and ϵ_{eskdcy}. Subsequently, σ^2_α and σ^2_ϵ are estimated by maximum likelihood. We use an Empirical Bayes (EB) procedure, also known as the Bayesian shrinkage estimator, to predict the random parameters α_d.[9]

2.3.1 Testing the Random Assignment

In a typical value-added estimation, sorting of students to teachers is a major threat to the identification of a causal parameter. High-achieving students might choose the best DTAs, leading to VA estimates that are upward biased. In the current setting, classes are randomly sorted using an algorithm.[10] This section tests the reliability of the random assignment of students to classes, which is necessary to validate the subsequent empirical analysis.[11]

[9]The Bayesian shrinkage estimates are a best linear unbiased predictor of each DTA's random component, which takes into account the variance (signal to noise) and the number of observations for each DTA. Specifically, estimates with a higher variance and a smaller number of observations are shrunk toward zero (Carrell and West (2010)).

[10]The university used the Myeconomics platform, which generates a stratified random assignment algorithm (using individual characteristics such as gender, and course of provenience) to place students into classes of equal size within each course/year.

[11]In principle, students are unable to change either classes or tutors. Students are reallocation in certain circumstances, with the permission of the undergraduate office. In our dataset, 7% of the students were observed in multiple classes within the same course. Nearly 40% of these "multiple

Testing the Random Allocation: Students' Achievement

Due to the scope of this study, we are particularly interested in testing the (lack of) selection associated with students' ability. The data provided by the University do not contain information about students' academic achievement prior to enrollment in the undergraduate program. We conduct two randomness tests on students attending 2^{nd} year courses for whom we can observe 1^{st} year marks. This applies to four core modules (Macroeconomics II, Microeconomics II, Economics II and, Econometrics) for three academic years (2009/10 to 2011/12).

Table 2.2: Testing the random allocation of students to classes (a VA approach)

	(1)	(2)	(3)
DTA value-added	0.00003670	0.00000003	0.00005283
	(0.00150350)	(0.00000124)	(0.00194919)
Observations	2,325	2,325	2,325
Number of DTAs	32	32	32
Number of classes	217	217	217
Course × Year FE	YES	YES	YES
Class-level controls	NO	YES	YES
Student-level controls	NO	NO	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. We focus on students attending 2^{nd} year courses, for which we can also observe 1^{st} year marks. This is the case in 4 core modules (Macroeconomics II, Microeconomics II, Economics II, and Econometrics) for three academic years (2009/10 to 2011/12). The dependent variable is the weighted average of all the final marks earned by the student in the 1^{st} year of the program. Class-level controls include: class size; dummy variables for the day/time of the week; shares of peers in the class that are female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. Column 1 includes Course × Year FE. Column 2 adds class-level controls. Column 3 adds student-level controls. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in section 2.3.

observations" were in a different class taught by the same tutor. It is not possible to distinguish the class assigned to the student and the class attended by the student in the data. For this reason, we exclude from the sample students who are not exclusively assigned to a class. However, we conduct robustness checks including these students in the sample and results are practically unchanged. This is discussed in detail in Section 2.4.

The first test estimates Equation (2.1). The dependent variable is the weighted average of all the final marks earned by the student in the 1^{st} year of the program. Final marks are a weighted average of all evaluations in each course. Under random assignment of students to classes, we expect DTAs to teach to a pool of 2^{nd} year students who have a similar distribution of ability. Therefore DTAs' value-added measures for 2^{nd} year courses, estimated using 1^{st} year marks, should not be statistically different from zero. The results are in Table 2.2, providing first empirical evidence in support of the random allocation hypothesis.[12]

The second test follows Carrell and West (2010). We test the random placement of students to classes using resampling techniques. For each class, we randomly draw (without replacement) 10,000 classes of equal size from the population of students attending the same course. For each simulated class, we calculate the average mark of the students in the 1^{st} year. We then compute p-values for each class, representing the fraction of simulated classes with the average mark smaller than the real one. Under random assignment, any p-value is equally likely to be observed, and therefore the expected distribution of the empirical p-values should be uniform. We test the uniformity of these distributions using the Kolomogorov-Smirnov test and a χ^2 test. The results are in Table 2.3. We reject the null hypothesis of uniform distribution for 1 of 12 course/year tests at the 5% level using the Kolmogorov-Smirnov test. We never reject the null hypothesis in 12 tests using the χ^2 test at the 5% level. These tests confirm the reliability of the random allocation procedure.

Testing the Random Allocation: Students' Covariates

We perform two supplementary tests on *all the sample*. Results are in appendix Tables B.1 and B.2. The first test is similar to Equation (2.1), but employs student-specific time invariant characteristics as an outcome. These are: (i) Gender, (ii) overseas status, and (iii) enrollment in the Department of Economics. The results are in appendix Table B.1. As we should expect, DTAs value-added measures are not statistically different from zero, indicating that no significant differences exist in these dimensions across groups of students enrolled in the same course, but allocated

[12] We did the same analysis using the mark earned by the students in the final exams of each course as a dependent variable. Results are qualitatively unchanged, and are available upon request.

Table 2.3: Testing the random allocation of students to classes (KS test and χ^2 test)

	Empirical p-values		Kolmogorov-Smirnov test	χ^2 test
	Mean	SD	(N failed / total tests)	(N failed / total tests)
Macroeconomics II	0.501	0.275	0/3	0/3
Microeconomics II	0.497	0.290	1/3	0/3
Economics II	0.480	0.276	0/3	0/3
Econometrics	0.496	0.306	0/3	0/3

Notes: the test focuses on students attending 2^{nd} year courses, for which 1^{st} year marks are also available. This is the case in 4 core modules (Macroeconomics II, Microeconomics II, Economics II and, Econometrics) for three academic years (2009/10 to 2011/12). For each class, we randomly draw (without replacement) 10,000 classes of equal size from the population of students attending the same course. For each simulated class, we calculate the average mark of the students in the 1^{st} year. We then compute p-values for each class, representing the fraction of simulated classes with the average mark smaller than the real one. We report the number of Kolmogorov-Smirnov and χ^2 tests for the uniformity of the distribution of p-values (that should be expected under random assignment of students to classes) failing at the 5% significance level.

to different DTAs.

The second test follows Braga et al. (2016). Using the three sets of students' characteristics as a dependent variable (gender, overseas status, and enrollment in the Department of Economics) we run separate Probit regressions for each course/year, including class-specific dummies. The null hypothesis is that the coefficients on the class dummies in each model are jointly equal to zero, which amounts to testing for the equality of the means of the observables, across classes within the same course. Appendix Table B.2 contains the results. We have 32 tests for each student characteristic, as we focus on eight courses in four years. At the 5% significance level, we never reject the null hypothesis for gender, we reject four times the null hypothesis for overseas status, and we reject twice the null hypothesis regarding being enrolled in the Economics Department. While we believe that the outcome of this test is very satisfactory, the estimation of the value-added model will be shown to be robust to the inclusion of individual and peer-level controls.[13]

[13]DTAs are also randomly allocated to seminar classes. This is to prevent any perception of unfairness in the teaching schedule for doctoral students. The random allocation of DTAs to classes allays any concern that value-added estimates might be contaminated by confounding factors. For instance, we might be worried that a DTA teaches at the preferred day and time of the week. We do not have personal information about DTAs, which constrains our ability to test the reliability of the random assignment of DTAs to seminar classes. However, the empirical estimation of the value-added model will be shown to be stable to the inclusion of indicators on the date/time of

2.4 Results

DTAs' value-added estimates obtained using Equation (2.1) are presented in Table 2.4. Column 1 shows the baseline regression, where Course $\times$ Year fixed effects are included. The resulting standard deviation in DTA quality, as measured by the value-added model, is around 0.085. This estimate is significant at the 1% level. Columns (2) to (4) test the robustness of the estimate to the inclusion of type of evaluation fixed effects, class-level controls, peer controls, and individual-level controls. Estimates are stable and precise across specifications, and always below the 1% significance level.[14] A 1σ change in DTA quality increases students' test scores by about 8.5% σ.[15]

Table 2.5 employs a unique feature of our dataset: information on marks for different evaluations - assessment, end-of-term test, and final exam. We exploit *within-course* variation *across type of evaluations* to search for heterogeneity in DTAs' value-added estimates. Arguably, at least two dimensions of variation might matter in this context: (i) Different evaluations carry a different weight in the final mark; and (ii) these evaluations occur at different times, while the DTA is teaching, and after the teaching component of the course has finished. Assessments taken during term time have the lowest weight in the final mark, with a mean of 6% (see Table 2.1). Tests are written examinations that take place right at the end of each term and carry a slightly higher weight (average of 8.6%). Exams occur at the end of the academic year usually after a long study break, and typically account for 80% of the final mark.

Table 2.5 presents DTAs' value-added estimates by type of evaluation obtained using Equation (2.1). DTAs have a large impact on assessments: a one standard deviation change in DTA quality increases students' test scores by around 0.16 of a standard deviation. VA estimates drop by around 65% for end-of-term tests, to 0.06

the class.

[14] Appendix Table B.3 tests the sensitivity of the precision of the main results to the change in the level of clustering. We report results unclustered, and clustered at the tutor and class level. The inference is practically unchanged.

[15] Appendix Table B.4 provides the DTAs' value-added estimates obtained using Equation (2.1), including in the sample students that are observed in more than one class within the same course (as described in Footnote 11). Results are practically unchanged.

Table 2.4: DTAs "value-added" model

	(1)	(2)	(3)	(4)
DTA value-added	0.0858***	0.0858***	0.0841***	0.0866***
	(0.0239)	(0.0239)	(0.0240)	(0.0260)
Observations	24,914	24,914	24,914	24,914
Number of DTAs	66	66	66	66
Number of classes	589	589	589	589
Course × Year FE	YES	YES	YES	YES
Evaluation FE	NO	YES	YES	YES
Class-level controls	NO	NO	YES	YES
Student-level controls	NO	NO	NO	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. The sample includes Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (from 2008/09 to 2011/12 included). Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. Column 1 includes Course × Year FE. Column 2 adds evaluation fixed-effects. Column 3 adds class-level controls. Column 4 adds student-level controls. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

of a standard deviation. The estimates for both assessment and end-of-term tests are significant at the 1% level. Finally, the DTAs' value-added estimates are very close to zero for final exams, implying that there is no variation in DTAs effectiveness in the final exams.[16]

Robustness Checks

The number of DTAs and the number of classes vary across columns in Table 2.5. While this is intrinsic to the set-up, we test the robustness of the findings focusing on a sample of courses that involve assessments, tests, and exams. This is the case for Microeconomics I and Econometrics. That is, each column will now show value-added estimates obtained on the *same* sample of courses, classes and tutors. Results are in appendix Table B.5. The value-added measure for assessment is now around

[16]The table shows a VA estimate and a related standard error of zero. These approximate 2.48×10^{-8} for the value-added, and 7.67×10^{-7} for the standard errors.

Table 2.5: DTAs "value-added" model by type of evaluation

	(1) Assessment	(2) Test	(3) Exam
DTA value-added	0.164***	0.0603**	0
	(0.0454)	(0.0268)	(0)
Observations	11,859	6,542	6,513
Number of DTAs	56	41	66
Number of classes	442	317	591
Course × Year FE	YES	YES	YES
Class-level controls	YES	YES	YES
Student-level controls	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. The sample includes Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (from 2008/09 to 2011/12 included). Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. Column 1 reports "value-added" estimates for assessment, column 2 for end-of-term tests, column 3 for final exams. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

0.10 of a standard deviation, significant at the 1% level.[17] Estimates for tests and exams are practically unchanged, suggesting that the decay in the estimates is not driven by a systematic change in the composition of the sample.

The estimates for assessments are around 0.16 of a standard deviation: this magnitude is larger than that for tests and exams. We test the robustness of this result, running separate regressions (only using assessments as an outcome) that exclude one academic course at a time. Results are in appendix Table B.6. We observe remarkably stable estimates and confidence intervals in each column of the table, indicating that the larger results for assessment are unlikely to be driven by outliers.

We also investigate whether our results are driven by differences in students' type across the different bands offered in the undergraduate program. We test the robust-

[17]The sample size is significantly reduced in this exercise. In the baseline regression of Table 2.5 we have 11,859 observations to estimate the value-added for assessment. In this table we use 6,148 observations.

ness of our results focusing only on the students enrolled in the Economics band, which is the largest one. Results are provided in appendix Table B.7. Column 1 reports value-added estimates using the pooled sample of evaluations, while columns 2 to 4 lay out the value-added estimates separately by type of evaluation. Point estimates barely change and results remain qualitatively unchanged in comparison to those reported in Tables 2.4 and 2.5.

Discussion

The decay in the estimates is a key finding. While previous research in post-secondary education has detected long-term teacher-effects for subsequent courses taught by a different teacher, and even in the labor market (Braga et al. (2016)), we are unaware of other studies detecting a sharp VA decay for subsequent evaluations of the same course, taught by the same tutor. We hypothesize that this decay could be connected to two non-exclusive forces (Jacob et al. (2010)): (i) Students may forget information or lose skills that they acquired under a particular teacher; (ii) students may make endogenous investments (that might be amplified by the importance of the evaluation) that could mitigate the consequences of the assignment to a particular teacher.[18] In particular, students could compensate for being assigned to a low-value-added DTA by exerting more effort to prepare for the final exam. Classes are only in the 1^{st} and 2^{nd} term and the final exam is in 3^{rd} term after a long period of self-guided study, which includes a month without no teaching. Moreover, this is a top UK Economics Department with high-achieving undergraduate students who may be well equipped to compensate for variance in DTA performance during the period of self-preparation.

These two aspects are closely connected and therefore difficult to disaggregate. However, the next section extends the empirical analysis to detect further informative heterogeneity of the VA estimates, while shedding light on the channels contributing to their decay.

It is important to highlight that our estimates are informative about potential

[18] End-of-term tests and final exams are always performed individually, within the university. Assessments instead are a more mixed type of evaluation. They could involve individual or group assessments taken at home or in class (such as individual or group presentations, for example). The data do not allow for a further inspection of the various types of assessments.

variation in DTAs' effectiveness. Nevertheless, our estimates do not provide information in term of *levels* (i.e., could be that *all* DTAs provide a substantial contribution on their students' academic performance, but there is no difference on this individual contribution *across* DTAs). Thus, policy implications of our results could differ depending on the goal. For example, if the Department wants heterogeneity on DTA's individual contributions to their students' performance in the intermediate assessments but not in the final exam (which carries the highest weight), our results deliver this outcome.

2.4.1 Heterogeneity in the Estimates and Other Results

This section examines DTAs' value-added estimates by students' year of study. Then it focuses on two student characteristics: overseas status and gender. Finally, it investigates the correlation between DTA quality and students' evaluation of the DTA.

Students' Year of Study

Students in UK universities must obtain 40 points out of 100 in each of the 1^{st} year courses to proceed to 2^{nd} year. Courses in the 1^{st} year do not contribute to students' grade at the end of the undergraduate program, which is a weighted average of 2^{nd} and 3^{rd} year marks.[19] This could be an informative source of variation. If the lack of persistence in the VA estimate is connected with students' endogenous responses to an increase in the weight of the evaluation, we would expect to observe a drop in the value-added estimates in the 2^{nd} year of study, when students are likely to exert more effort.

Results by students' year of study are shown in Columns 1 and 2 of Table 2.6. DTAs' value-added estimates are 0.12σ for courses taken in the 1^{st} year of study and 0.03σ for courses taken in the 2^{nd} year of study. The estimate for 1^{st} year courses is significant at the 1% level, while the estimate for 2^{nd} year courses is not statistically different from zero. The results are consistent with the hypothesis that when the evaluations are more important students exert more effort, reducing the

[19] Courses in the 3^{rd} year are not in the sample, as these are optional modules and do not have a DTA.

contribution of the DTA to the outcome. However, 2^{nd} year students might be more mature, experienced, and willing to work more (independently of the importance of the exam). Overall, both these dynamics (weight of the test and maturity of the student) could reduce the extent to which DTAs affect students' academic success.

The composition of the evaluations in 1^{st} and 2^{nd} year courses is rather different.[20] This might affect the interpretation of the above estimates. That is, the drop in the VA measures in the 2^{nd} year might be driven by the change in composition of the evaluations in the sample of analysis, rather than by a change in students' personal investments, such as an increase in effort. We investigate this by estimating the two separate samples (years 1 and 2) by type of evaluation. Figure 2.1 reports estimates and 95% confidence intervals obtained using Equation (2.1). The estimates for assessment are around 0.2σ in the 1^{st} year, and 0.1σ in the 2^{nd} year. A similar decay in the estimates (steeper in 1^{st} year course) is observed across types of evaluation. To conclude, these exercises show that VA estimates are larger for 1^{st} year courses and for assessments (both in years 1 and 2). The findings appear to be consistent with the hypothesis that students work harder on assessments that carry a higher weight, potentially mitigating the consequences of the assignment to a particular teacher. This would reduce the VA coefficient.

Students' Demographics

Students' demographics represent another interesting source of variation. We have information about gender and overseas status for all the students in the sample. One might posit that female students would exert more effort, since male high-school students are overwhelmingly less likely than female high-school students to spend time doing homework or other study set by teachers (OECD (2015)). Overseas students might exert more effort because they pay higher fees at UK universities.[21]

[20] In the sample of 1^{st} year courses, we have 5,691 assessments, 2,185 tests and 3,040 exams. These account for 52%, 20% and 28% respectively of the evaluations in 1^{st} year courses. In the sample of 2^{nd} year courses, we have 6,168 assessments, 4,357 tests and 3,473 exams. These account for 44%, 31% and 25% of the 2^{nd} year course evaluations.

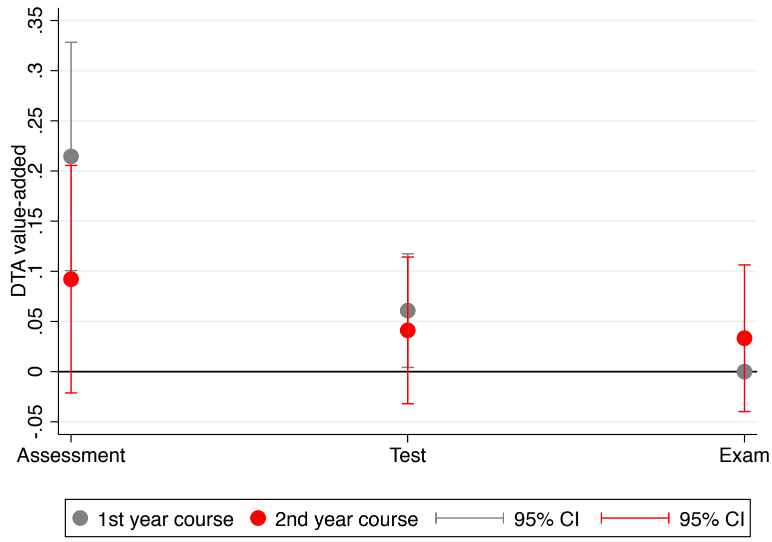
[21] UK and EU students pay about £9,200 per year. Overseas students pay about £17,460 per year. See here.

Table 2.6: DTAs "value-added" model (further heterogeneity)

	Year of the Program		Gender		Overseas Status	
	(1)	(2)	(3)	(4)	(5)	(6)
	1^{st} year course	2^{nd} year course	Female	Male	Overseas	Home/EU
DTA value-added	0.126***	0.0280	0.0881***	0.0746***	0.0973**	0.123***
	(0.0316)	(0.0174)	(0.0295)	(0.0199)	(0.0463)	(0.0191)
Observations	10,916	13,998	9,616	15,298	10,587	14,327
Number of DTAs	31	43	66	66	66	66
Number of classes	289	300	572	588	582	582
Course × Year FE	YES	YES	YES	YES	YES	YES
Evaluation FE	YES	YES	YES	YES	YES	YES
Class-level controls	YES	YES	YES	YES	YES	YES
Student-level controls	YES	YES	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. The sample includes Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (from 2008/09 to 2011/12). Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. Columns 1 and 2 report the results for 1^{st} year and 2^{nd} year courses, respectively. Columns 3 and 4 show the results for female and male, respectively. Columns 5 and 6 report the results for overseas and home/EU students, respectively. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Figure 2.1: DTA "value-added" by type of evaluation and year of the undergraduate program



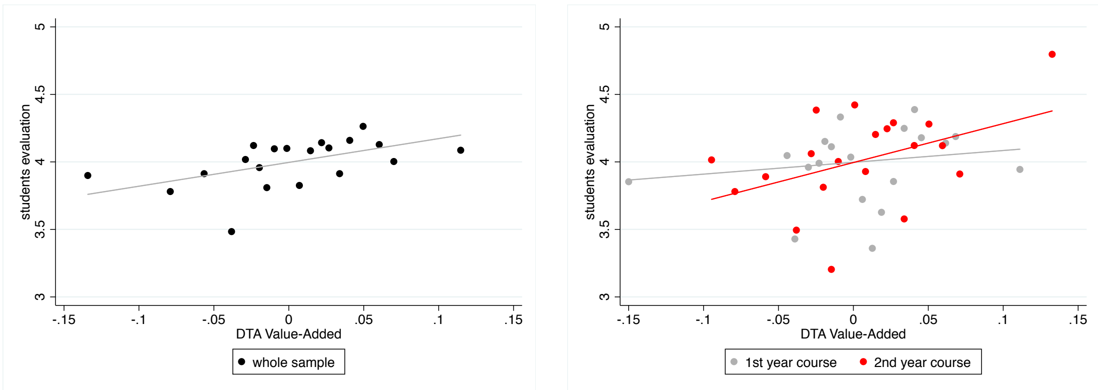
Notes: The table shows DTA "value-added" by type of evaluation and year of the undergraduate program. Standard errors are clustered at the DTA level. The sample includes Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (from 2008/09 to 2011/12). Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

We estimate a VA model for these two samples. Columns 3 and 4 of Table 2.6 show the results by gender. Columns 5 and 6 of Table 2.6 show the results by overseas status. Results are quite balanced by students' gender: we detect a value-added estimate for males of 0.07σ, and for females of 0.09σ. Both estimates are significant at the 1% level. Results are also relatively similar for overseas students (0.10σ) and for home-EU students (0.12σ). Both the coefficients are precisely estimated under the conventional significance levels. This part of the analysis while leading to balanced precise estimates, does not improve our understanding of the mechanisms behind the lack of persistence of the VA estimates.

Students Evaluations of the DTAs

Figure 2.2 shows the correlation between DTAs value-added estimates and the students' evaluations of the DTAs. We regress the students' evaluations of each DTA in each course on their predicted value-added, controlling for Course × Year fixed effects. The correlation is positive and statistically significant, and is stronger for 2^{nd} year courses.[22] Although we should be cautious not to over-interpret this difference, its direction seems to suggest that 2^{nd} year students are more precise in evaluating DTA effectiveness than 1^{st} year students. Also, 2^{nd} year students might be paying more attention to the teachers because 2^{nd} year courses carry higher weight.

Figure 2.2: DTA "value-added" vs. students evaluation of the DTA



Notes: This table shows the correlation between DTA "value-added" vs. students evaluation of the DTA. Left panel presents a scatter plot using the whole sample of analysis. Right panel presents two scatter plots, one for each year of the undergraduate program. The fitted lines correspond to a regression of the students evaluation score on the estimated DTA value-added, which also includes Course × Year FE.

2.5 Concluding Remarks

The value-added approach is the most prominent method for evaluating teacher quality. While studies using value-added models indicate that teacher quality is an important determinant of student achievement in primary and secondary education (Kane and Staiger (2008), Rothstein (2010b), Chetty et al. (2014b)), there is less evidence for how the quality of instruction affects student outcomes at the post-

[22]Results in table format are omitted for brevity considerations only, and are available upon request.

secondary level. This is because the exams are less likely to be standardized and students and teachers are typically endogenously matched on the basis of their ability (Carrell and West (2010), Braga et al. (2016)).

This paper studies a widely-employed but overlooked resource in post-secondary education: doctoral teaching assistants. We argue the role and influence of DTAs on student achievement is likely to be growing over time (Lusher et al. (2015)), and is worth studying. We focus on a UK Economics undergraduate program. The sample advances the research because: (i) Students are randomly assigned to classes; (ii) we have access to a detailed individual database, which allows the first estimation of doctoral teaching assistants' value-added models.

We validate the empirical approach using a variety of tests to confirm the random allocation of students to classes. We find that a one standard deviation change in DTA quality increases students' academic achievement by around 8.5 percent of a standard deviation. A novel estimation of the within-course dynamics shows a startling decay: Value-added estimates are larger for assessments taken during the course, drop for end-of-term tests and are not statistically different from zero for final exams, implying that there is no variation in DTAs effectiveness for the final exams. Various robustness checks confirm that the lack of persistence of the VA estimates is genuine and it is not driven by changes in the sample composition, nor by outliers.

While previous research in post-secondary education has detected long-term teacher-effects for subsequent courses (taught by a different teacher) and even in the labor market (Braga et al. (2016)), in this study we find a sharp VA decay for distinct evaluations within the same course taught by the same teacher. In general, the persistence of the teacher effects could reflect two main mechanisms (Jacob et al. (2010)): (i) Students may make other investments that mitigate or exacerbate the consequences of the assignment to a particular teacher; and (ii) students may forget information or lose skills that they acquired under a particular teacher.

The analysis suggests that the short-term decay in the estimates might be driven by students' endogenous investments (mechanism 1). For example, students could compensate for being assigned to a low-value-added DTA by working harder in the self-guided study period. This increase in effort could be amplified by the high weight of the final exam in the final mark. We also acknowledge that even in our short time-frame, the lack of variation in DTAs' effectiveness detected in final exams can be due to students losing skills that they acquired under a particular teacher (mechanism 2). Finally, we believe that DTAs have very little power, experience, or incentive to change their way of teaching, causing such dramatic changes in the VA estimates across separate types of evaluations.[23]

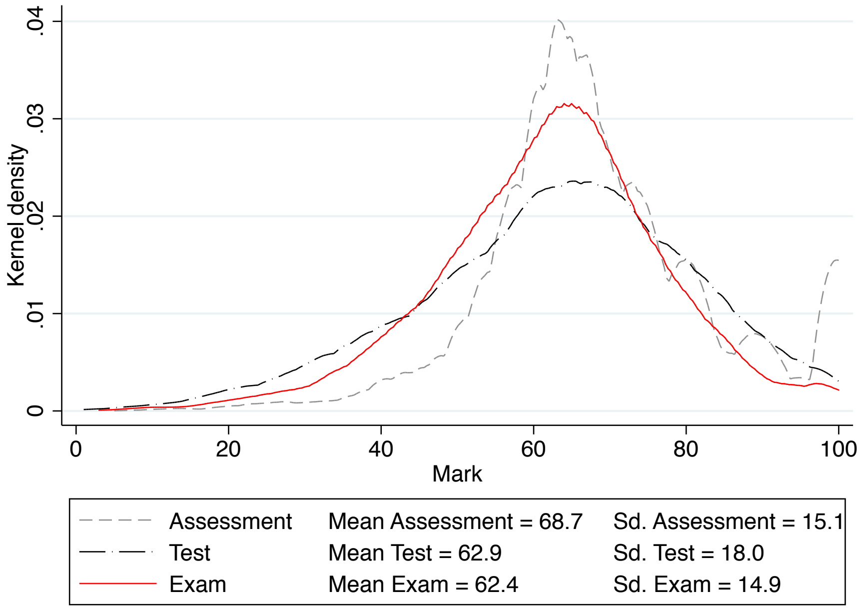
Further research is warranted to better understand the sensitivity of value-added estimates to a different set of incentives affecting students' responses. Also, the student sample is highly selected which may limit the external validity of this study. We encourage more research into the interactions between students' initial level of ability, personal investments and sensitivity of the value-added estimates. Very little is known in this area, providing promising avenues for future research.

[23]Carrell and West (2010) present evidence that professors who excel at promoting contemporaneous student achievement teach in ways that improve their student evaluations but harm the follow-on achievement of their students in more advanced classes.

Appendices

Figures

Figure B.1: Distribution of students' marks by type of evaluation



Tables

Table B.1: Testing the random allocation (a VA approach on all sample of analysis)

	(1) Female	(2) Overseas	(3) Economics
DTA value-added	0.0102285	0	0
	(0.0289321)	(0)	(0.1075076)
Observations	6,649	6,649	6,649
Number of DTAs	66	66	66
Number of classes	589	589	589
Course × Year FE	YES	YES	YES
Class-level controls	YES	YES	YES
Student-level controls	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. Outcome variables are gender (column 1), overseas status (column 2) and enrolment in the Economics Department (column 3). All outcomes are dummy variables. All the specifications include Course × Year FE, class-level controls and individual-level controls. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Table B.2: Testing the random allocation (χ^2 tests on all sample of analysis)

	Female		Overseas		Economics	
	$P - value < 5\%$	$P - value < 1\%$	$P - value < 5\%$	$P - value < 1\%$	$P - value < 5\%$	$P - value < 1\%$
	(N failed / total tests)		(N failed / total tests)		(N failed / total tests)	
The World Economy	0/4	0/4	0/4	0/4	1/4	1/4
Economics I	0/4	0/4	0/4	0/4	0/4	0/4
Macroeconomics I	0/4	0/4	0/4	0/4	0/4	0/4
Microeconomics I	0/4	0/4	1/4	0/4	0/4	0/4
Macroeconomics II	0/4	0/4	1/4	0/4	0/4	0/4
Microeconomics II	0/4	0/4	1/4	0/4	0/4	0/4
Economics II	0/4	0/4	0/4	0/4	1/4	0/4
Econometrics	0/4	0/4	1/4	1/4	0/4	0/4

Notes: In this test we run separate Probit regressions for each course/year, including class-specific dummies. Standard errors are clustered at the DTA level. Outcome variables are gender (columns 1/2), overseas status (column 3/4) and enrolment in the Economics Department (column 5/6). The null hypothesis is that the coefficients on the class dummies in each model are jointly equal to zero. We have 32 tests for each student characteristic, as we focus on 8 courses in 4 years. The test statistics are χ^2, with varying parameters depending on the number of classes in each course/year. We report results at the 5% and at the 1% significance level.

Table B.3: Robustness check clustering standard errors at different levels

	(1)	(2)	(3)	(4)
DTA value-added	0.0858	0.0858	0.0841	0.0866
SE (clustered tutor level)	(0.0239)***	(0.0239)***	(0.0240)***	(0.0260)***
SE (non-clustered)	(0.0139)***	(0.0139)***	(0.0139)***	(0.0144)***
SE (clustered student level)	(0.0131)***	(0.0131)***	(0.0133)***	(0.0136)***
SE (clustered class level)	(0.0136)***	(0.0136)***	(0.0139)***	(0.0144)***
Observations	24,914	24,914	24,914	24,914
Number of DTAs	66	66	66	66
Number of classes	589	589	589	589
Course × Year FE	YES	YES	YES	YES
Evaluation FE	NO	YES	YES	YES
Class-level controls	NO	NO	YES	YES
Student-level controls	NO	NO	NO	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors are reported in parenthesis. Standard errors clustered at student level and class level were computed using a bootstrap procedure. Bootstrap Standard errors were computed using 1,000 bootstrap samples. The sample includes Macroeconomics 1, Microeconomics 1, Economics 1, The World Economy, Macroeconomics 2, Microeconomics 2, Economics 2, and Econometrics (from 2008/09 to 2011/12 included). Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, overseas status, and enrolment in the Department of Economics. Column 1 includes Course × Year FE. Column 2 adds evaluation fixed-effects. Column 3 adds class-level controls. Column 4 adds student-level controls. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Table B.4: Robustness check including students observed in different classes

	(1)	(2)	(3)	(4)
DTA value-added	0.0854***	0.0854***	0.0833***	0.0867***
	(0.0228)	(0.0228)	(0.0228)	(0.0245)
Observations	26,767	26,767	26,767	26,767
Number of DTAs	66	66	66	66
Number of classes	597	597	597	597
Course × Year FE	YES	YES	YES	YES
Evaluation FE	NO	YES	YES	YES
Class-level controls	NO	NO	YES	YES
Student-level controls	NO	NO	NO	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. The sample of analysis includes students that are observed in multiple classes within the same course. Given that there is no information regarding the initial allocation of these students, nor of the actual class attended, we randomly selected one observation per student at course/year level. That is, each student with multiple observations in different classes (within the same course) is randomly allocated to only one class. Column 1 includes Course × Year FE. We then add sequentially evaluation fixed effects, class-level controls, and student-level controls, as in Table 2.4. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Table B.5: Robustness check on the sample including same DTAs and classes

	(1) Assessment	(2) Test	(3) Exam
DTA value-added	0.0985***	0.0607***	0
	(0.0283)	(0.0217)	(0)
Observations	6,148	4,126	1,928
Number of DTAs	27	27	27
Number of classes	168	168	168
Course × Year FE	YES	YES	YES
Class-level controls	YES	YES	YES
Student-level controls	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. The sample includes courses that have assessments, end-of-term tests, and exams. These are Microeconomics I and Econometrics. Focusing on these 2 courses ensures that VA estimates by type of evaluation are obtained on a sample that includes the same tutors and classes. All the specifications include Course × Year FE, class-level controls, and student-level controls. Column 1 shows the results for assessment. Column 2 for end-of-term tests. Column 3 for final exams. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Table B.6: Robustness check excluding one course at the time (assessments only)

	(1)	(2)	(3)	(4)	(5)	(6)
DTA value-added	0.130***	0.159***	0.160***	0.160***	0.164***	0.207***
	(0.0339)	(0.0439)	(0.0599)	(0.0444)	(0.0520)	(0.0475)
Observations	10,156	10,699	9,031	10,173	10,697	8,539
Number of DTAs	49	51	48	47	51	39
Number of classes	363	379	365	366	380	347
Course $\times$ Year FE	YES	YES	YES	YES	YES	YES
Class-level controls	YES	YES	YES	YES	YES	YES
Student-level controls	YES	YES	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. The sample includes assessments only. The courses that have assessments are: The World Economy, Economics I, Microeconomics I, Macroeconomics II, Economics II and Econometrics. All the specifications include Course $\times$ Year FE, class-level controls, and student-level controls. Column 1 shows the results excluding The World Economy from the sample. We then exclude sequentially: Economics I, Microeconomics I, Macroeconomics II, Economics II. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Table B.7: Robustness check on the sample of students enrolled in the Economics band of the undergraduate program (L100)

	(1) Pooled	(2) Assessment	(3) Test	(4) Exam
DTA value-added	0.107***	0.192***	0.0711**	7.60e-08
	(0.0172)	(0.0298)	(0.0290)	(2.30e-07)
Observations	16,197	7,113	4,934	4,150
Number of DTAs	57	46	40	57
Number of classes	460	316	305	460
Course × Year FE	YES	YES	YES	YES
Evaluation FE	YES	N/A	N/A	N/A
Class-level controls	YES	YES	YES	YES
Student-level controls	YES	YES	YES	YES

Notes: ***, ** , * Denote statistical significance at 1%, 5% and 10% level respectively. Standard errors reported in parenthesis are clustered at the DTA level. The sample of analysis includes only students that are enrolled in the Economics band of the undergraduate program. Marks are standardised within type of evaluation/course/year to have a mean of zero and a standard deviation of one. Class-level controls include class size, dummy variables for the day/time of the week, shares of peers in the class that were female, with an overseas status, and who were enrolled in the Department of Economics. Student-level controls include dummies for: gender, and overseas status. Column 1 reports "value-added" estimates for the pooled sample of evaluations, column 2 only for assessment, column 3 only for end-of-term tests, and column 4 only for final exams. The procedure to obtain the "value-added" estimate and related-confidence interval is explained in Section 2.3.

Chapter 3

Points To Save Lives: The Effects of Traffic Enforcement Policies on Road Fatalities

3.1 Introduction

Certain public policies aim to influence individual behavior to discourage harmful activities such as crime, tax evasion or traffic violations. In this paper, we study the effects of a change in traffic enforcement policies for two reasons. First, reducing the incidence of traffic accidents is a first-order public health issue in most countries. According to recent estimates from the World Health Organization, more than one million people die in car accidents each year worldwide (WHO (2013)), generating an economic cost of approximately 1.5% of global GDP. Hence, better enforcement policies have the potential to yield large economic and social gains at a potentially low cost. A second reason to study the impact of traffic enforcement policies is that the key outcome variables in this setting, such as the number of accidents or road fatalities, are directly observable and collected in a standardized way across many countries. This is in contrast with tax evasion or crime, which are only imperfectly observed and therefore much harder to quantify.

In recent decades, many countries have implemented penalty points systems (PPS) to administer driving licences. Under a PPS, drivers are allocated a fixed number of points that they can lose if they commit traffic violations. Losing all points can result in the suspension of the driving licence for a period of time, or its permanent withdrawal.[1] According to the theoretical analysis of Bourgeon and Picard (2007), the PPS contributes to reducing the gap between the private and social valuation of the cost of traffic accidents, which the judicial system and the insurance market fail to address. By doing this, the PPS achieves two social objectives. First, it screens reckless drivers to ensure that they lose their licence. Second, it acts as a deterrence mechanism for normal drivers. Their analysis also shows that complementing this nonmonetary sanction system with moderate fines, as is usually done is most countries, is likely to increase overall welfare.

We study the effects of the introduction of a penalty points system in Spain in 2006 on road fatalities. Since the reform was introduced at the national level, it is challenging to find a suitable counterfactual to estimate the effect on drivers' behavior and mortality rates. Traditionally, case studies have chosen a set of "reasonably" similar units (e.g., cities or States) to form a counterfactual (for example, Card (1990), Card and Krueger (1994)). A problem with these case studies is that the choice of the counterfactual group is to some extent arbitrary.

We use the synthetic control method developed by Abadie and Gardeazábal (2003) and Abadie et al. (2010). The basic idea of this method is to build a counterfactual using a weighted average of all potential control units for which data is available. The set of potential controls is called the "donor pool". Each of the units in the donor pool is assigned a nonnegative weight in the synthetic group and, by construction, these weights must add up to one. Following this procedure, the method makes explicit the contribution of each comparison unit to the counterfactual of interest. Having constructed the counterfactual, measuring the impact of the policy simply requires comparing the evolution of road fatalities in Spain to the same aggregate variable for the *synthetic* Spain.

The main advantage of the synthetic control method over traditional case studies is the rigorous way in which the control group is chosen. By requiring only the use of pre-intervention data, the method is "blind" about the impact of the choices made to select the counterfactual on the final estimates. Moreover, the restriction that

[1]In some countries, the policy is called "Demerit Points System" (DPS), where drivers start out with zero points and they accumulate them with traffic offenses. In that case, the licence is withdrawn when a certain number of points is reached.

the weights of each unit must be nonnegative and add up to one implies that there is no use of extrapolation to build the counterfactual, something that often occurs in linear regression analysis (sometimes inadvertently). Under certain assumptions that we describe in more detail in Section 3.3, the synthetic control method provides a compelling identification strategy to evaluate the impact of the PPS in Spain.

To build the synthetic control group, we use data on several predictors of road fatalities for all EU-15 countries, excluding Spain.[2] Our newly-constructed dataset includes annual data on alcohol consumption, road density, GDP per capita, fuel consumption, and the stock of vehicles, among other variables, for the period 1990-2011. The main source of data is the World Road Statistics report, produced by the International Road Federation.[3] We describe the dataset in more detail in Section 3.2.2.

Our main finding is that annual road fatalities in Spain declined substantially between 2006 and 2011 as a result of the introduction of the PPS. By 2011, there were 4.4 road fatalities per 100,000 people in Spain, compared to 7.2 in the synthetic control group (a 39% difference). We estimate the cumulative reduction in road fatalities over that period to be approximately 20%, a very substantial effect. This is an interesting result, given that previous research has found short-run effects of traffic enforcement policies on drivers behavior, but these effects start to decline rapidly (Abouk and Adams (2013)). Following the implementation of the PPS, it was a sharply increase in the number of driving licenses suspended by the judiciary authority. This evidence might suggests that the persistent reduction in road fatalities could not only be driven by deterring a risky-driving behavior, but also by taking reckless drivers out of the roads.

We conduct two placebo exercises to check the robustness of this result. First, we evaluate the estimated effect under the assumption that the PPS was adopted

[2]The term EU-15 refers to the members of the European Union during the period 1995-2004, namely: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and the United Kingdom.

[3]The International Road Federation is a non-profit organization based in Switzerland focused on "promoting the development and maintenance of better, safer and more sustainable roads and road networks". In consultation with national statistical institutes, it collects comprehensive data on road networks, traffic and inland transport all over the world. For more information, see www.irfnews.org.

by Spain in the year 2000 (six years before the actual implementation). We find no effect of this placebo policy on road deaths in Spain. As a second placebo test, we estimate the effect of a hypothetical adoption of the PPS in 2006 in each of the countries in the donor pool. We would expect the gap between road fatalities in each of these countries and their synthetic controls to be zero on average. Indeed, we find that the negative effect on road fatalities estimated for Spain stands out as the largest of all. These two placebo tests strongly suggest that our main findings are not due to pure chance.

Finally, we use our year-by-year estimates of the number of lives saved due to the PPS to calculate the economic benefits derived from this policy. During the five-year period after its adoption, the PPS prevented approximately 3,500-4,000 road fatalities (depending on the donor pool used in the estimation). According to Abellán et al. (2009), the value of a statistical life (VSL) in Spain is estimated to be in the vicinity of €1.3 million ($1.82 million). Therefore, we estimate that the PPS yielded an economic benefit of €4.6-5.1 billion ($6.0-6.7 billion) in its first five years of implementation, which corresponds to 0.43-0.48% of Spain's GDP.

Previous Literature

Economists have long been interested in the effects of traffic enforcement policies on individual behavior. Since the pioneering work of Peltzman (1975), researchers have analyzed the effects of mandatory seat belt laws (Evans (1986), Cohen and Einav (2003)), minimum legal drinking age (Lovenheim and Slemrod (2010)), minimum wage policies and alcohol-related traffic accidents (Adams et al. (2012)), banning text messaging while driving (Abouk and Adams (2013)), fines and experience-rated insurance premiums (Dionne et al. (2011)), motorcycle helmet mandates (Dee (2009)), smoking bans in bars (Adams and Cotti (2008)), and the regulation of bar's opening hours (Green et al. (2014)), among others.

Some studies have focused specifically on the effects of penalty points system in different countries. DePaola et al. (2013) study the effects of the PPS in Italy, using a regression discontinuity design in which the assignment variable is time. They compare the number of road accidents and fatalities just before and after the introduction of the new law, estimating a 9% fall in traffic accidents and a 18-30%

reduction in road fatalities. Several studies have attempted to evaluate the effects of the penalty points system (PPS) in Spain. These studies use monthly traffic statistics[4] and employ a variety of time-series techniques to address the issue of seasonality in the road fatalities data, such as controling nonlinearly for long-run trends and other variables associated to traffic accidents. Castillo-Manzano et al. (2010) estimate that the PPS led to a 12.6% reduction in the number of deaths on highway accidents (not weighted by population). Novoa et al. (2010) find a 10% reduction in the risk of death or serious injury in highway accidents for *drivers*. Pulido et al. (2010) estimate a 14.5% reduction in the number of road fatalities in the 18 months following the implementation of the PPS. Even though the outcome variable is slightly different in each study, the general conclusion from this set of studies is that the PPS reduced mortality in traffic accidents by 10-14% in the 18-24 months after the introduction of the policy.

The time-series methods used in these early analyses of the Spanish PPS have several limitations. First, they may not control adequately for long-run trends in road fatalities that are unrelated to the introduction of the PPS, for example a change in the pace of improvement in vehicles' safety features or road quality. During the decade starting in 2000, there was a generalized decline in the number of road fatalities across all countries in the EU-15. Failing to account for this trend could introduce a negative bias in the estimated effects of the Spanish PPS on fatalities, i.e. overestimating the effects of the policy. Even though we do not observe the factors behind these trends directly, we effectively control for them by using other EU-15 countries (likely affected by similar shocks as Spain) to build the synthetic control group. Second, controling for seasonality in road fatalities is difficult unless the time series used is very long, but the studies mentioned above focus on the period between January 2000 and December 2007, which is not long enough to accurately capture the seasonal swings in road fatalities. In this study, we use annual frequency data so we avoid having to deal with the varying seasonal patterns across EU-15 countries (e.g., due to differences in weather conditions and holiday schedules).

The key contribution of this paper to the literature is to use the synthetic control

[4] Specifically, the monthly data on traffic accidents and fatalities in Spain is published by the General Directorate of Traffic (*Dirección General de Tráfico*, DGT).

method, which provides a more reliable estimate of the medium-term effect of the penalty points system in Spain. A second contribution of the paper is the cost-benefit evaluation of the PPS, using estimates of the value of a statistical life. In the Spanish case, the cost-benefit analysis indicates that the PPS yielded substantial economic as well as human benefits with a very low cost of implementation (DGT (2011)), suggesting that countries without such a system should consider adopting it in the future.

The rest of the paper is organized as follows. Section 3.2 describes the institutional background and the dataset. Section 3.3 summarizes the synthetic control method. Section 3.4 presents the main results. Section 3.5 provides some additional discussion of the results, including the cost-benefit analysis using estimates of the value of a statistical life (VSL). Section 3.6 concludes.

3.2 Institutional Background and Data

3.2.1 The Spanish Penalty Points System (PPS)

The penalty points system (PPS) reform was enacted by the Spanish parliament on July 19th, 2005 and became effective on July 1st, 2006. The reform was widely discussed in the media, received almost unanimous support from all political parties, and was publicized through an extensive information campaign in the media (DGT (2011)).

In the Spanish PPS, drivers start out with 12 points (eight for newly-licensed drivers) that they can lose if they commit traffic violations. Some examples are speeding violations (two to six points depending on the severity of the offense), drunk driving (four to six points) or using the mobile phone while driving (three points). If a driver loses all points, their driving licence is suspended for a period of six months. Until 2011, 107,000 drivers had lost their licence, corresponding to 0.4% of all existing licences (DGT (2011)). Penalized drivers can recover their licence after going through traffic safety workshops organized by the traffic authorities. Losing all points for a second time may result in the permanent withdrawal of the driving licence.

Official government reports state that the main goal of the policy was to modify driving behavior through non-monetary penalties, seen as a necessary complement to the existing (and essentially unchanged) regime of monetary penalties (DGT (2011)). Additionally, they expressed the intention of shifting to individual drivers the responsibility of retaining their licence through good behavior.

Several other EU-15 countries have implemented similar points programs, as shown in Table 3.1. West Germany was the first country to implement such a system in 1974, followed by France in 1992, Greece in 1993, and the United Kingdom in 1995. As we explain below, we exclude some countries where the PPS was introduced around the same time as in Spain fromt the donor pool to ensure that they do not bias the results.

Table 3.1: Penalty Points Systems by Year of Adoption

Country	Year of Adoption	Type of System
Germany (West)	1974	Demerit
France	1992	Demerit
Greece	1993	Gain
United Kingdom	1995	Gain
Ireland	2001	Gain
Luxembourg	2002	Demerit
Italy	2003	Demerit
Netherlands	2003	Gain
Austria	2005	Take into account recidivism
Denmark	2005	Gain
Spain	2006	Demerit
Belgium	-	-
Finland	-	-
Portugal	-	-
Sweden	-	-

Source: European Transport Safety Council (2011).

3.2.2 Data

We construct a new dataset of traffic-related statistics for EU-15 countries using multiple data sources, for the period 1990-2011. For data on traffic fatalities,[5] road network quality, the stock of four- and two-wheeled vehicles and other characteristics

[5]Fatalities are measured under the internationally standardized "30-day" measure, which counts any deaths within the first 30 days after the accident.

at the country level, we use the World Road Statistics (WRS), a report published by the International Road Federation.[6] For all these variables, we calculate the outcomes per 100,000 people, using population from the World Bank Open Data. We also use GDP figures in US dollars from the same source. Finally, we collect data on alcohol consumption per capita, a strong predictor of road accidents, from the OECD iLibrary.

The final dataset has some limitations. Even though the IRF makes an effort to collect homogeneous data across countries, there is a substantial proportion of missing values for certain variables that would have been good predictors of road accidents. For example, incomplete information about the annual volume of road traffic in each country prevents us from using this variable in the set of predictors of road fatalities. For some other variables, the number of missing values is limited to some specific country-year observations. In all cases where there are no more than two consecutive missing values, we use linear extrapolation to fill the gaps in the data so that we can include those variables in the analysis.

Another limitation is the relatively small sample size, as we have data for 15 countries over 22 years ($N = 330$). This is not uncommon in the synthetic control literature. For example, in the seminal paper by Abadie and Gardeazábal (2003), the sample size consisted of 17 Spanish regions, and the synthetic group included only two regions. As we explain below in more detail, this is less critical in this setting than in a linear regression context. The reason is that we are using aggregate data, so there is no statistical uncertainty about the representativeness of the sample, as noted by Abadie et al. (2010). Therefore, we do not construct confidence intervals around the point estimates in same way that we would when using a sample of data. However, there is a different source of uncertainty regarding the precision of the estimates and their statistical significance, which depends on the validity of the sythetic control group. As we show below, the results are robust to a number of placebo checks.

Finally, our data has annual frequency instead of monthly. Even though this may seem to be a limitation because it reduces the effective sample size, notice that annual data does not feature seasonality, removing a potential source of bias from

[6]See Foonote 2 above.

the estimation. For this reason, we argue that the advantages of monthly data are somewhat overstated in the earlier studies on this topic reviewed above. Having 15 years of complete pre-intervention data allows us to construct the counterfactual for Spain using the synthetic control method.

3.3 Synthetic Control Method

In this section we provide a formal description of the synthetic control method. For a more comprehensive treatment, please see Abadie et al. (2010).

Assume we observe $J+1$ countries over T periods, and let country 1 (in this case, Spain) receive an intervention in period $t_I < T$, so that this country is exposed to the treatment during periods $t_I+1, ..., T$. Let Y_{it} be the value of the outcome for country i in year t, and let Y_{it}^N be the value that would have been observed in the absence of the intervention. The goal is to estimate the effect of the intervention on the treated country, that is $\alpha_{1t} = Y_{1t} - Y_{1t}^N$, for the post-intervention periods $t \in (t_I+1, ..., T)$. Since Y_{1t}^N is not observed in the post-intervention period, we need to estimate a counterfactual.

The key idea of the synthetic control method is to construct this counterfactual using a linear combination of the potential control units. For this purpose, we define the vector $W' = (w_2, ..., w_{J+1})$, which contains the weights that will be assigned to each unit in the donor pool. We impose that these weights must be nonnegative and sum up to one:

$$w_j \geq 0, \quad \forall\, j \in (2, ..., J+1) \tag{3.1}$$

$$\sum_{j=2}^{J+1} w_j = 1 \tag{3.2}$$

These two conditions guarantee that the comparison group is constructed without using extrapolation, as emphasized by Abadie et al. (2010). The optimal vector of weights is chosen to minimize the distance between the characteristics of the synthetic control and the treated country. Let X_1 be a $(k \times 1)$ vector of pre-intervention characteristics of the treated country. Similarly, define the $(k \times j)$ matrix X_0, which contains the same variables for the untreated countries. Thus, the vector $W^{*'} = (w^*_2, ..., w^*_{J+1})$ is chosen to minimize:

$$\min \quad ||X_1 - X_0W||_V = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)} \quad \text{s.t. eqs. (3.1) and (3.2)} \tag{3.3}$$

where V is a $(k \times k))$ symmetric and positive semidefinite matrix. In general, V is a diagonal matrix with main diagonal elements $(v_1, ..., v_k)$, where v_l represents the relative weight assigned to the l-th pre-intervention variable when measuring the distance between the treated country and the synthetic group. Thus, if X_{jl} represents the value of the l-th pre-intervention variable for country j, the optimal weights w^*_2, ..., w^*_{J+1} minimize:[7]

$$\sum_{l=1}^{k} v_l \, (X_{1l} - \sum_{j=2}^{J+1} w_j X_{jl})^2 \tag{3.4}$$

In other words, the weights are chosen to minimize the difference in pre-intervention characteristics between the synthetic control and the treated country. Under the assumption that a large-enough period of pre-intervention data is available (i.e., $t_I \gg 0$), an approximately unbiased estimator for the parameter of interest α_{1t} is given by:

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w^*_j Y_{jt} \quad for \quad t = (t_{I+1}, ..., T) \tag{3.5}$$

We can also express the cumulative effect in the k years after the intervention as:

$$\sum_{t=t_I}^{t_{I+k}} \widehat{\alpha}_{1t} = \sum_{t=t_I}^{t_{I+k}} \left(Y_{1t} - \sum_{j=2}^{J+1} w^*_j Y_{jt} \right) \tag{3.6}$$

[7]In order to ensure that the counterfactual resembles the treated country, both in terms of the pre-treatment characteristics and the outcome of interest, Abadie et al. (2010) suggest the use of lagged values of the outcome variable in the process of obtaining the optimal weights.

3.4 Results

We implement the estimation routine outlined above using the `synth` command for Stata, developed by Abadie et al. (2010).[8]

3.4.1 Composition of the Synthetic Control Group

We consider two alternative donor pools: i) All EU-15 countries excluding Spain, ii) and the subsample of countries that do not have a penalty points system (PPS) or adopted it before the year 1996. The reason for restricting the donor pool in ii) is that the subset of countries that adopted the PPS around the same time as Spain could potentially contaminate the synthetic group, because we expect the policy to have an effect on their road fatalities. We keep in the data countries that had adopted the PPS before 1996 because it is unlikely that the policy still has a relevant effect on *changes* in road fatalities more than ten years after its introduction (although we would expect it to have an impact the *level* of that outcome).

Table 3.2 reports the weights assigned to each country under the two alternative donor pools. The first column shows the weight when all countries are included in the donor pool. The country with the highest weight in the *synthetic* Spain is Portugal (0.35). This is not surprising, given that Portugal is a neighboring country that shares many geographic and cultural characteristics with Spain. Other countries with substantial weights are Belgium (0.25), Sweden (0.17), Italy (0.16), Greece (0.07) and Luxembourg (0.006).

The second column of Table 3.2 reports the weights when the donor pool is restricted to avoid contamination from the introduction of similar penalty points systems in other countries. This reduces the donor sample to eight countries, and it changes significantly the composition of the *synthetic* Spain. France, another country that shares a border with Spain, receives a high weight (0.30), although the highest goes to Belgium (0.42). Portugal's weight reduces to only (0.02), while Finland (0.18) and Greece (0.09) also receive a positive weight.

[8] More information about the command for Stata and other software platforms is available here.

Table 3.2: Country weights for Synthetic Spain

	Donor Pool	
Country	All countries	No PPS (or adopted before 1996)
Austria	0	-
Belgium	0.245	0.415
Denmark	0	-
Finland	0	0.176
France	0	0.300
Germany	0	0
Greece	0.070	0.093
Ireland	0	-
Italy	0.162	-
Luxembourg	0.006	-
Netherlands	0	-
Portugal	0.351	0.016
Sweden	0.166	0
United Kingdom	0	0

Notes: This table reports the weights assigned to each country using the synthetic control method explained in Section 3.3. By construction, the weights are nonnegative and must add up to one. In the first column, the pool of donors contains all EU-15 countries excluding Spain. In the second column, the pool of donors contains the subset of countries that have never adopted a penalty points system (PPS) or adopted it before the year 1996.

Table 3.3 reports summary statistics for all the predictors of road deaths in the pre-intervention period (1990-2005). We also report the average of the main outcome variable, road deaths per 100,000 people, for the years 1995, 2000 and 2005. The first column reports the actual values for Spain, the second for the synthetic Spain constructed using all potential control countries, the third for the synthetic Spain using only the uncontaminated subset, and the fourth shows the (unweighted) averages for all 14 potential donors.

The two synthetic controls do a reasonably good job at matching the values of the predictors of the *true* Spain. This is partly due to the fact that EU-15 countries are fairly homogeneous on many socioeconomic characteristics. However, the limited size of the donor pools implies that some of the values do not match perfectly. The differences between Spain and synthetic Spain are below 10% (which we consider a good match) for log GDP per capita, alcohol consumption per capita, kilometers of total and secondary roads, number of four-wheel vehicles and diesel consumption.

The match is less precise for the number of motorcycles, petrol consumption and road density. For these three variables, Spain has lower values than the two synthetic controls. Despite these differences, the average values of the main outcome of interest–road deaths per 100,000 people–are matched almost perfectly at three different years in the pre-intervention period (1995, 2000 and 2005). This provides reassurance that the synthetic control method is performing quite well in both cases, despite the sample size limitations.

3.4.2 Main Results

Figure 3.1 shows that the two alternative synthetic control groups track the evolution of road fatalities per 100,000 people in Spain for the entire pre-intervention period. The top panel compares Spain against the synthetic Spain constructed using all other EU-15 countries. The two trend lines follow each other closely in the pre-intervention period (1990-2005), featuring a pronounced decline of road deaths per 100,000 people from about 17.9 in 1990 to 9.4 in 2005. After the PPS was introduced in Spain (in July 2006), the country experienced an even further decline in road deaths down to 4.4 fatalities per 100,000 people in the year 2011, compared to a value of 7.2 for the *synthetic* Spain.

The bottom panel of Figure 3.1 shows the trends for Spain and the second *synthetic* Spain, constructed using the restricted donor pool. As in the top panel, the pre-intervention trends in road deaths are very similar. The evidence of an effect of the PPS starting in the year 2006 is clear, and the divergence between Spain and the synthetic control by the year 2011 is essentially the same, as the alternative synthetic Spain has 7.1 deaths per 100,000 people.

These results imply that the PPS led to a large reduction in road fatalities in Spain during the first five years of application of the penalty points system. Table 3.4 reports the evolution of road fatalities per 100,000 people in Spain and the two alternative synthetic control groups for the period 2007-2011. By 2011, the incidence of road fatalities in Spain was 39% lower than that of *synthetic* Spain (4.41 vs. 7.23). The percentage difference is very similar regardless of the donor pool used.

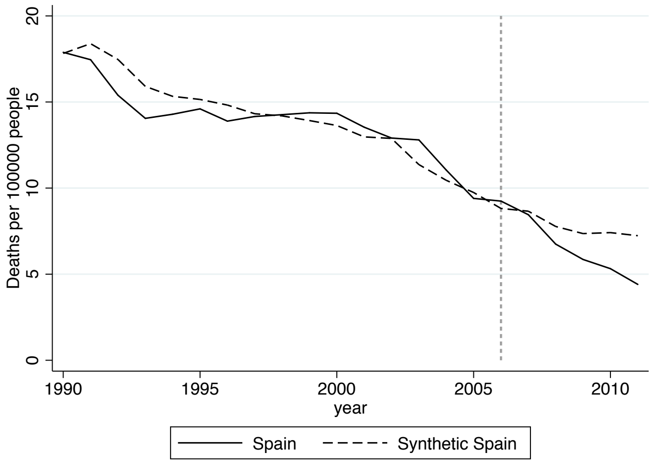
Table 3.3: Summary Statistics: Spain vs. Synthetic Spain

	Spain	Synthetic Spain: All countries	Synthetic Spain: No PPS (or adopted before 1996)	Average of all potential donors
Log GDP per capita (USD)	9.71	10.08	9.86	10.13
Alcohol consumption per capita (lts.)	11.34	11.60	10.47	11.24
Total roads (km. per 100,000)	1,211.45	1,427.75	1,158.04	1,283.57
Secondary roads (km. per 100,000)	344.30	311.73	325.79	313.46
Road density, (sq. km. per 100,000)	0.002	0.022	0.018	0.049
Four-wheel vehicles (per 100,000)	48,009.24	47,910.01	45,699.48	46,941.76
Motorcycles (per 100,000)	4,453.86	4,790.38	5,603.44	4,727.66
Diesel consumption (lts. per 100,000)	43.31	41.20	37.79	49.60
Petrol consumption (lts. per 100,000)	20.83	26.68	29.24	44.13
Road deaths, 2005 (per 100,000 people)	9.40	9.67	9.74	8.44
Road deaths, 2000 (per 100,000 people)	14.35	13.05	13.63	11.38
Road deaths, 1995 (per 100,000 people)	14.60	13.77	15.15	12.58
Road deaths, 1990 (per 100,000 people)	17.88	17.99	17.82	14.89

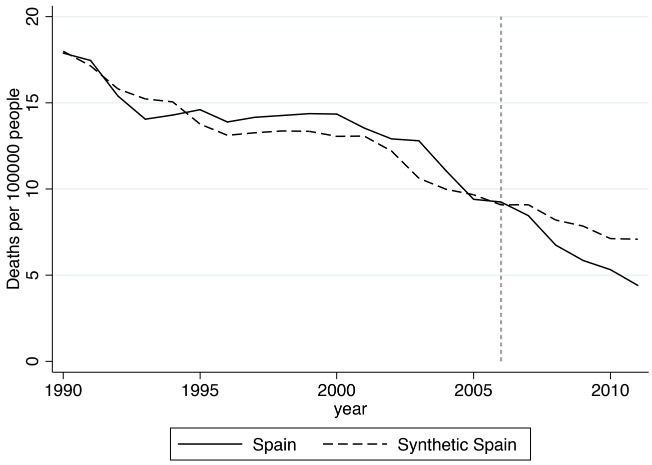
Notes: All variables are averaged for the 1990-2005 period, except road deaths which is evaluated in three specific years. GDP per capita is measured in US dollars. Alcohol consumption is measured in litres per person/year. Total and secondary roads are measured in kilometers per 1,000 people. Road density is measured in squared km. per 100,000 people. The number of four-wheel vehicles and motorcycles is measured in units per 100,000 people. Diesel and petrol consumption are measured in litres per 100,000 people and year. Road deaths are measured under the standardized 30-day measure, which counts any deaths due to traffic accidents in the 30 days after the accident. In the second column, the pool of donors contains all EU-15 countries excluding Spain. In the third column, the pool of donors contains the subset of countries that do not have a penalty points system (PPS) or adopted the policy before 1996. The fourth column shows the (unweighted) averages for all 14 potential donors.

Figure 3.1: Road Deaths over Time: Spain vs. "synthetic" Spain

Donor Pool: All Countries



Donor Pool: No PPS, or adopted before 1996



Notes: To construct the top panel, we construct the synthetic control using all EU-15 countries excluding Spain. For the bottom panel, we use only the 8 countries without a penalty points system (PPS) and those that adopted it before 1996. The vertical dotted line indicates that the policy was enacted the year 2006.

If we focus on the cumulative difference over the first five years of implementation of the PPS, the difference is approximately 20% (30.78 vs. 38.43). These are substantial point estimates of the medium-term impact of the policy. The cumulative impact is almost twice as large as that estimated by earlier studies (discussed in the introduction), which focused on the short-term impact 18-24 months immediately after the introduction of the PPS. The last two columns of Table 3.4 report the annual difference in road fatalities per 100,000 between Spain and the two synthetic controls. We use these estimates of the number of lives saved due to the PPS to calculate the economic benefits of the policy in Section 3.5.

Table 3.4: Road Fatalities after PPS Adoption: Spain vs. "synthetic" Spain

		Synthetic Spain		Difference Spain vs. Synthetic Spain	
Year	Spain	All countries	No PPS (or adopted before 1996)	All countries	No PPS (or adopted before 1996)
2007	8.45	8.66	9.09	-0.20	-0.63
2008	6.75	7.77	8.20	-1.02	-1.45
2009	5.85	7.36	7.85	-1.50	-1.99
2010	5.32	7.41	7.12	-2.09	-1.80
2011	4.41	7.23	7.09	-2.83	-2.68
2007-2011	30.78	38.43	39.34	-7.65	-8.56

Notes: Road deaths are measured per 100,000 people under the standardized 30-day measure, which counts any deaths due to traffic accidents in the 30 days after the accident. "All Countries" indicates that the donor pool contains all EU-15 countries excluding Spain. "No PPS (or adopted before 1996)" indicates that the donor pool contains the subset of countries that do not have a penalty points system (PPS) or adopted the policy before 1996, i.e. at least ten years before Spain.

Following the implementation of the PPS, it was a sharply increase in the number of driving licenses suspended by the judiciary authority. Appendix Figure C.1 shows the number of driving licenses suspended between 2000 and 2009, adjusted by population size. Appendix Table C.1 lays out the number of driving licenses suspended during the same period separately by the length of the suspension.[9] After 2006, the number of suspensions suddenly rose from approximately 77 to 121 driving licenses per 100,000 people, representing a 63.63% increment in 4 years. This evidence might suggest that the persistent reduction in road fatalities could not only be driven by the impact of the PPS on deterring a risky-driving behavior, but also by its effect on taking reckless drivers out of the roads.

3.4.3 Placebo Tests and Statistical Significance

We now turn to analyzing the robustness of the results using a series of placebo tests as suggested by Abadie et al. (2010). First, we test whether we would find an effect by assuming that the PPS was introduced in Spain in the year 2000, six years before its actual introduction. Appendix Figure C.2 shows the trends for Spain

[9]This information is provided by the Spanish Directorate General for Traffic (DGT (2009)).

and synthetic Spain using the two alternative donor pools, as before. Recall that the method uses pre-intervention data to match the behavior of the main outcome variable in the treated country with the synthetic control.[10] In both cases, the match before 2000 is reasonably good. If the penalty points system had actually been passed in 2000, we would have expected the two lines to start diverging right after that year. However, in the two panels of appendix Figure C.2 we observe that both synthetic controls continue following the trend for Spain quite closely until 2006, when the policy was actually enacted. After that year, the two lines start diverging significantly, with the *true* Spain featuring a steeper downward trend after 2006. This suggests that the large effect obtained for the 2006 reform was not due to chance.

Second, we compare the gap in road fatalities in Spain vs. synthetic Spain against the gaps that we obtain by assuming that the policy had been introduced in each country in the donor pool exactly in the same year, 2006. For this exercise, we only use the subset of countries that do not have a PPS or adopted it before 1996. Hence, appendix Figure C.3 shows the gap between road deaths per 100,000 people in each country vs. its synthetic control, assuming that a penalty points system (PPS) was enacted in the year 2006 (marked by the vertical dotted line). Since the policy was only enacted in that year for Spain, for the other eight countries shown in the figure we would expect the gap to be zero even after 2006. Despite the fact that there are large swings in the gap for two of the countries,[11] by far the largest negative gap in 2011 corresponds to Spain (marked with the red solid line in the graph).

Following Abadie et al. (2010), we interpret this second placebo exercise to be *prima facie* evidence that the estimated effect of the policy is statistically significant. The argument is that, of all possible countries to which we could have applied the same estimation strategy, the effect found for Spain is the largest. Since there were nine countries (Spain plus the eight countries in the restricted donor pool), the statistical significance in this case is approximately 11% (= 1/9).

[10]This implies that the composition of the synthetic control changes (and therefore, the weights assigned to each unit in the donor pool), given that we moved the implementation date of the policy. Appendix Table C.2 reports the sets of weights used in this placebo test.

[11]The country with a large positive gap is Greece, the one with a large negative gap is Portugal.

3.4.4 Synthetic Control Group Weights and Potential Confounders

Table 3.2 reports the weights assigned to each country under the two alternative donor pools. There is an important difference in the weight received by Portugal across these two groups. To study if our results are driven by changes in the composition of the donor pool, we re-estimate the effect of the PPS by excluding from the donor pool one country at a time. Country weights for the fourteen synthetic Spain resulted from this exercise are reported in appendix Table C.3, while the comparison of these synthetic control groups against Spain is presented in appendix Figure C.4. Even though in some cases there is variation in the weight assigned to a country across the different synthetic control groups, the permanent decline in road fatalities after the implementation of the PPS is present in all specifications.

The persistent reduction in road fatalities could also been explained by other factors changing at the same time as the PPS, confounding the effect of this policy on road fatalities. For example, changes in the expenditure on new roads construction could be a potential confounder. An increase in the number of roads could reduce traffic density, diminishing the probability of traffic accidents. In addition, an increase in the expenditure on road maintenance could also reduce traffic accidents, due to an improvement of the roads. Unfortunately, we do not have information on road expenditure for all EU-15 countries, and therefore we cannot use this information in the construction of our synthetic counterfactual. Nevertheless, we do have this information for Spain, which allows us to study potential changes in trends that coincide with the implementation of the PPS. Appendix Figure C.5 shows the road expenditure in Spain on new construction and maintenance. Road expenditure on new construction does not follow a clear trend over the whole period, while expenditure on road maintenance shows an upward trend that starts much before the implementation of the PPS. Moreover, expenditure in both new road construction and road maintenance experience a decline from 2009 onwards, due to the global economic recession. This reduction on road expenditure should likely increase the number of road fatalities. Instead, we observe a reduction of road fatalities during this period. Hence, this evidence suggests that changes in road expenditure is not a relevant mechanism behind our results.

Changes in the Spanish economic activity could be another potential confounder of the effect of the PPS on road fatalities (Adams and Cotti (2008), Adams et al. (2012)). Even though we already included the GDP per capita as one of the pre-intervention variables in the construction of our synthetic counterfactual, we analize any potential change in its trend that could coincide with the implementation of the PPS. Appendix Figure C.6 shows the evolution of the GDP per capita in Spain. Since 2001 up to 2008, there is a remarkable upward trend. Nevertheless, the Spanish GDP per capita starts to decline in 2009 due to the global economic recession. If this drop in the Spanish economic activity had been a main driver of the reduction in road fatalities, we wouldn't have observed a declaine in road fatalities during 2006, 2007 and 2008.

3.5 Discussion of Economic Benefits of the Policy

Our estimate of a 20% reduction in road fatalities in the five years after the adoption of the PPS in Spain is substantially larger than earlier estimates (Castillo-Manzano et al. (2010), Novoa et al. (2010), Pulido et al. (2010)), which found reductions in road mortality in the vicinity of 10-15%. As discussed in the introduction, these studies were not able to adequately control for other factors that may have affected road safety over time in all EU countries simultaneously, such as the pace of improvements in vehicle safety technology. By providing a more compelling counterfactual, the synthetic control method removes the attenuation bias present in previous studies. It is also worth noting that these studies focus on the short-run effects of the policy, since they only use data for the two years after its adoption. In this study, we use data until 2011, which allows us to focus on the medium-term effects of the policy. This difference in the length of the evaluation period could also explains the discrepancy between previous estimates and our findings. Evidence regarding driving license suspensions suggests that the persistent reduction in road fatalities could be partly driven by taking reckless drivers out of the roads, which might explains even larger effects in the reduction in road fatalities in later years.

In order to do a cost-benefit analysis of the PPS, we separately estimate the benefits and costs of the policy for society as a whole. On the benefit side, we use estimates of the value of a statistical life (VSL) to obtain an economic equivalent of the number of lives saved. The value of a statistical life is the amount of money that individuals or societies are willing to spend to save a human life. This concept has a long history in the economics literature, and there are multiple estimates for many countries (see Viscusi (1993), Viscusi and Aldy (2003), Orley Ashenfelter (2004), Ashenfelter (2006)). Estimates of the VSL are used extensively by governments to evaluate public policies related to health and other risks.

In Spain, the Directorate General for Traffic (DGT) commissioned a report to estimate the VSL in order to perform cost-benefit analyses of their policies (Abellán et al. (2009)). The central estimates of the VSL obtained by this team of researchers is €1.3 million per life ($1.69 million at the 2011 exchange rate).

Multiplying these values by the number of lives saved each year between 2007 and 2011 (for the complete donor pool), we obtain a total benefit of €4.6 billion ($6 billion) over this five-year period. This is equivalent to 0,43% of the annual Spanish GDP. The figures for the restricted donor pool are slightly larger: 5.1 billion ($6.7 billion), which is equivalent to 0,48% of GDP. These economic benefits are close to 0.1% of GDP each year. The figures suggest an extremely high return of this policy in economic and human terms.

Even though there are no official estimates of the cost of implementing the penalty points system in Spain, the available information suggests that the cost was an order of magnitude smaller than the estimated benefits. A report by the Directorate General of Traffic (DGT (2011)) explains that the information campaign at the time of introduction of the policy received a substantial amount of free publicity in the media (especially on TV and the written press), which allowed the DGT to make large savings on advertising. Moreover, the traffic re-education courses through which where offending drivers can recover their points charge tuition fees, so the net cost of these courses for the authorities is also limited.[12] Therefore, we argue that the economic benefits estimated above can be considered a good approximation of the net benefit of the penalty points system.

[12]One could also argue that the costs for sanctioned drivers of attending the courses or having their license withdrawn should be included in the overall welfare calculation, but these costs are likely to be dwarfed by the benefits to society of having fewer accidents.

3.6 Concluding Remarks

This paper has studied the impact of the introduction of a penatly points system in Spain on mortality from traffic accidents. We have found a large and significant reduction in road fatalities by 20% over a five-year period, about twice as large as the estimates from earlier studies. Evidence suggests that the persistent reduction in road fatalities might not only be driven by the success of the PPS in deterring a risky-driving behavior, but also by taking reckless drivers out of the roads.

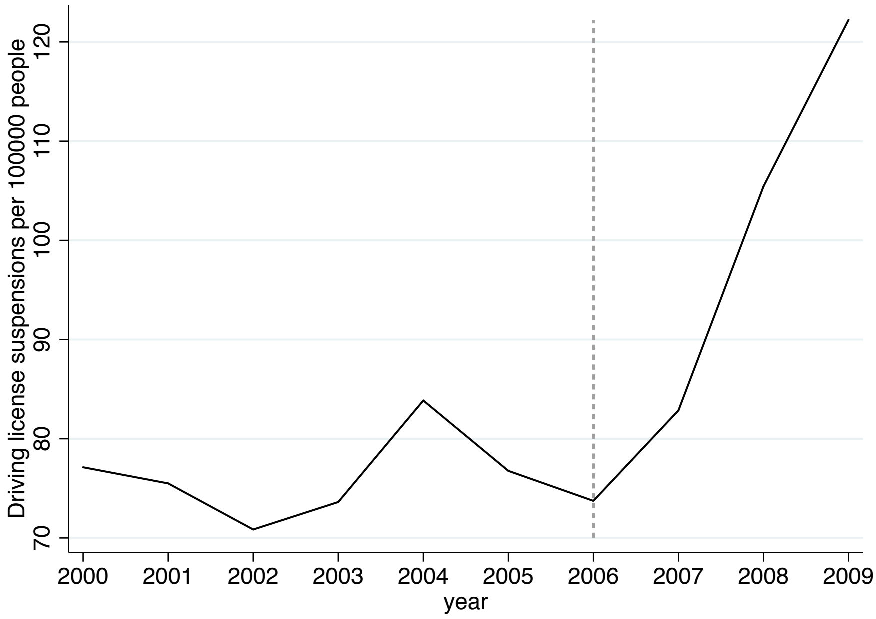
The results suggest that other countries, especially those currently in the middle and low-income categories, should consider adopting similar policies in order to reduce the mortality on their roads. Of course, this conclusion must come with the caveat that the context also matters: Spain is a highly-developed country with modern infrastructure and government institutions capable of managing sophisticed information systems to monitor millions of drivers.

More research is needed to understand the specific mechanisms that made the penalty points system so effective in Spain. Researchers could exploit some variation in traffic enforcement across regions in order to identify the relative importance of these mechanisms. For this purpose, it would be necessary to have access to high-frequency individual-level data on traffic violations, e.g. speeding, seat belt use, or alcohol consumption, which are not currently in the public domain.

Appendices

Figures

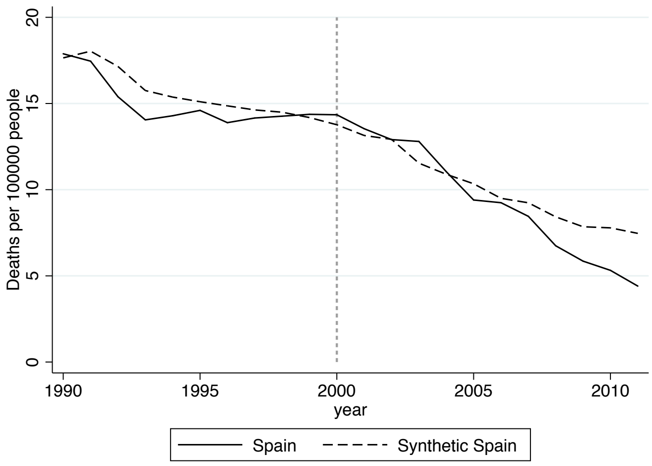
Figure C.1: Driving Licenses Suspended by the Judiciary Authority



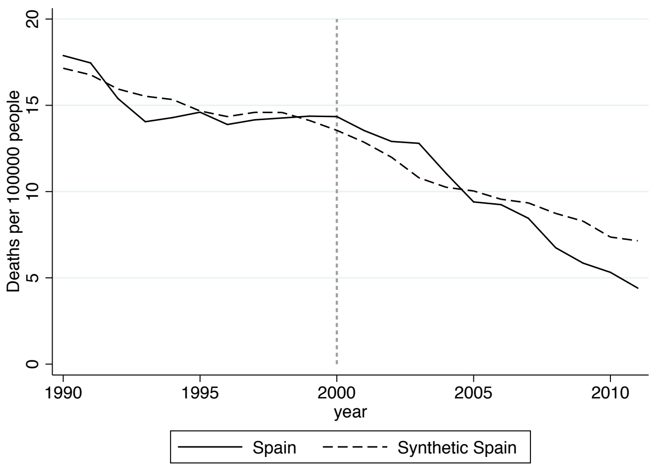
Source: Spanish Directorate General for Traffic (2009).

Figure C.2: Placebo Test: Moving the Policy to 2000

Donor Pool: All Countries

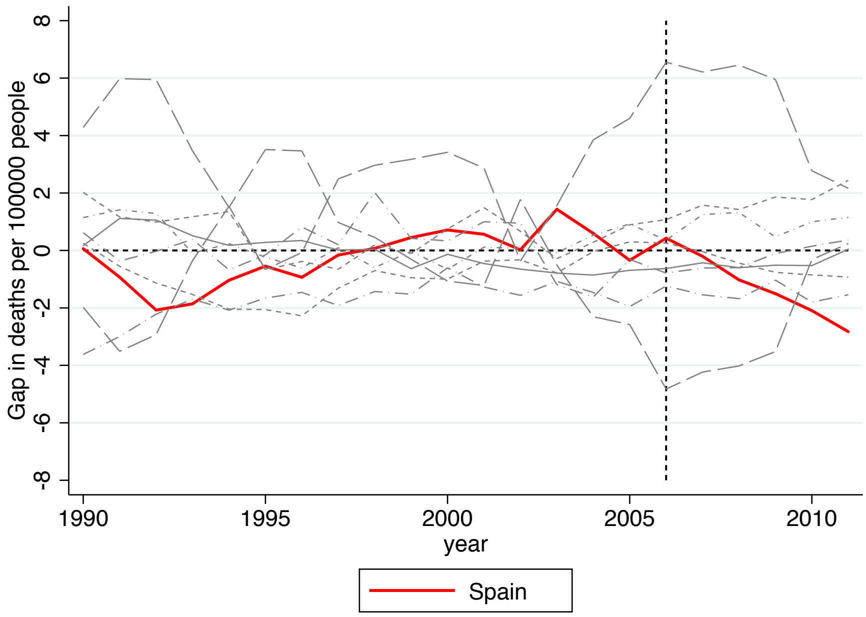


Donor Pool: No PPS, or adopted before 1996



Notes: In the top figure, we construct the synthetic control using all EU-15 countries excluding Spain. For the bottom panel, we use only the 8 countries without a penalty points system (PPS) and those that adopted it before 1996. The vertical dotted line indicates that the policy is assumed to have been enacted the year 2000 in Spain for this placebo test.

Figure C.3: Effect of PPS in Spain vs. Placebo Effects in Other Countries



Notes: This figure shows the gap between road deaths per 100,000 people in a country vs. its synthetic control, assuming that a penalty points system (PPS) was enacted in the year 2006 (marked by the vertical dotted line). The policy was only enacted in that year for Spain, whereas the other eight countries shown in the figure either do not have a penalty points system (PPS) or adopted it before 1996. For this countries, this is just a placebo exercise and we would expect the gap to be zero.

Figure C.4: Road Deaths over Time: Spain vs. “synthetic” Spain (excluding from the Donor Pool one Country at the Time)

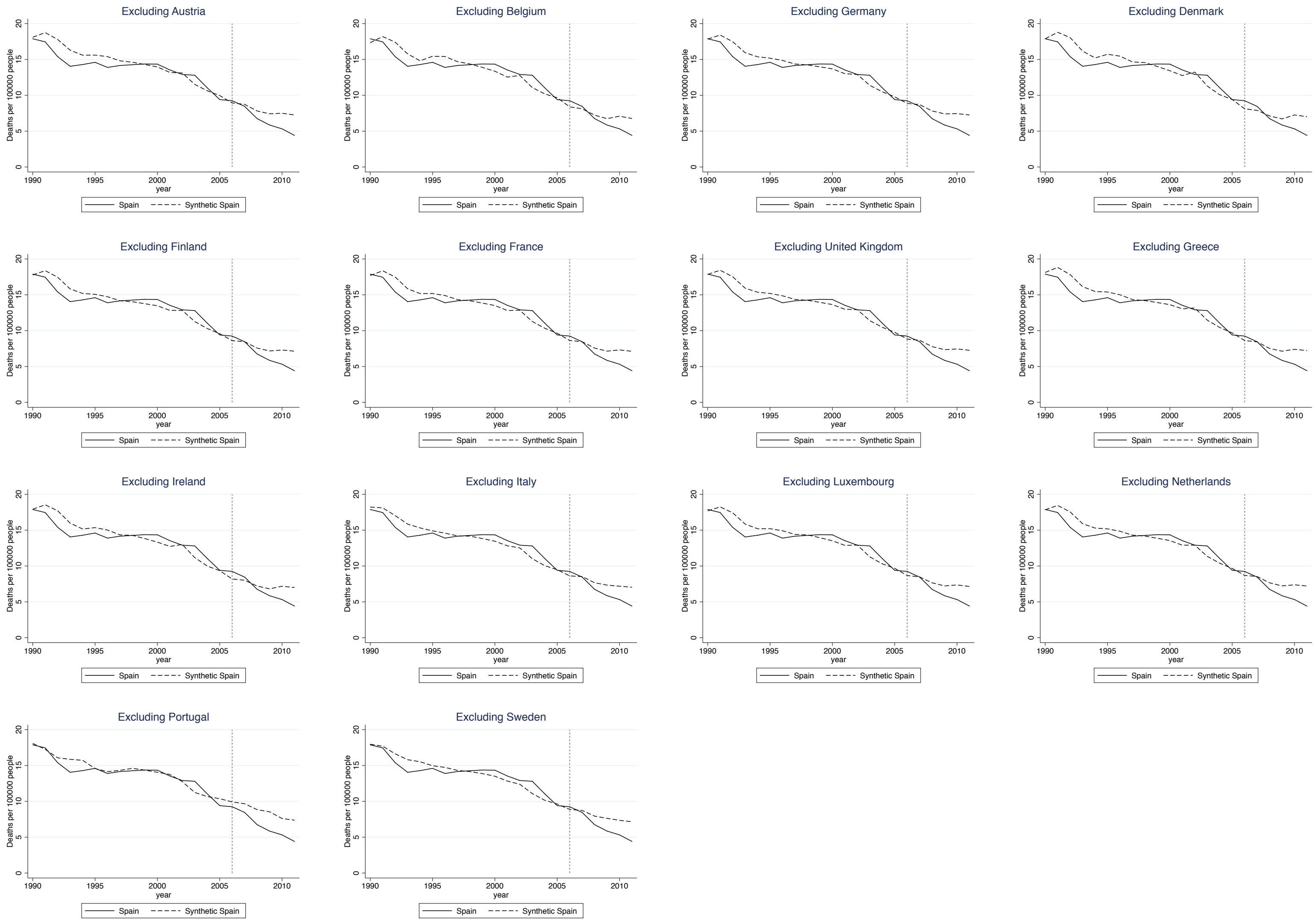
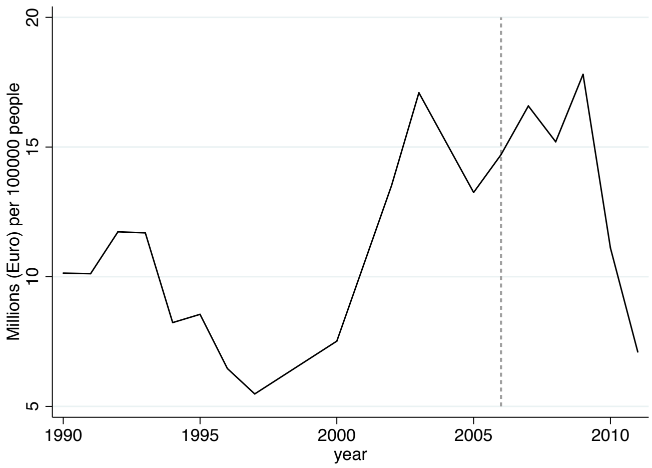
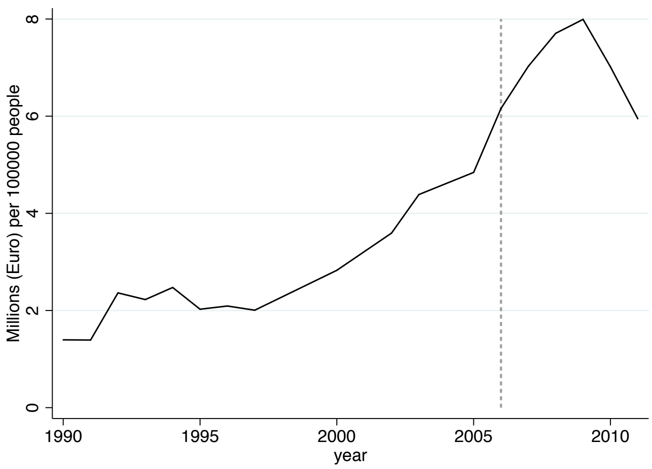


Figure C.5: Road Expenditure in Spain

Road expenditure in new construction

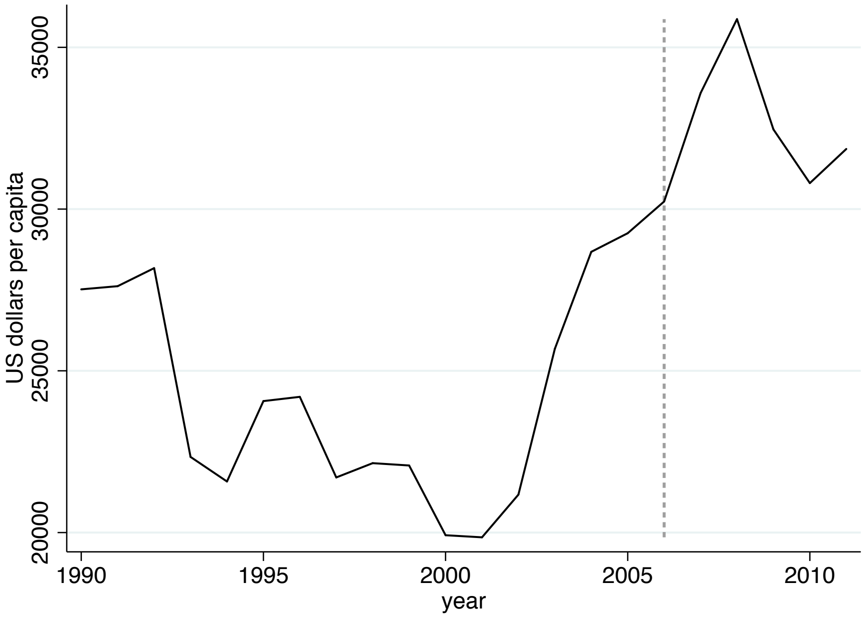


Road expenditure in maintenance



Notes: The top figure shows the evolution of road expenditure in new construction in Spain. The bottom figure shows the evolution of road expenditure in maintenance in Spain.

Figure C.6: GDP per capita in Spain



Tables

Table C.1: Licenses Suspended by the Judiciary Authority

	Length of Suspension				
Year	< 6 months	6 months - 1 year	1 year - 3 years	> 3 years	Total
2000	5,049	449	25,413	142	31,053
2001	3,511	280	26,826	153	30,770
2002	2,531	188	26,475	160	29,354
2003	1,837	2,888	26,164	170	31,059
2004	1,248	9,461	25,046	237	35,992
2005	817	11,176	21,305	211	33.509
2006	657	13,207	18,654	219	32,737
2007	632	16,627	19,902	314	37,475
2008	707	24,030	23,397	323	48,457
2009	675	29,199	26,356	440	56,670

Source: Spanish Directorate General for Traffic (2009).

Table C.2: Country weights for Synthetic Spain: Placebo Test

Country	Donor Pool: All countries	Donor Pool: No PPS (or adopted before 1996)
Austria	0	-
Belgium	0.193	0.051
Denmark	0	-
Finland	0.073	0.047
France	0	0.323
Germany	0	0.260
Greece	0.157	0.319
Ireland	0.009	-
Italy	0.197	-
Luxembourg	0	-
Netherlands	0	-
Portugal	0.282	0
Sweden	0	0
United Kingdom	0.089	0

Notes: This table reports the weights assigned to each country using the synthetic control method explained in Section 3.3. In this case, we conduct a placebo test on which we move the policy to year 2000. By construction, the weights are nonnegative and must add up to one. In the first column, the pool of donors contains all EU-15 countries excluding Spain. In the second column, the pool of donors contains the subset of countries that have never adopted a penalty points system (PPS) or adopted it before the year 1996.

Table C.3: Country weights for Synthetic Spain: Excluding from the Donor Pool one Country at the Time

	PANEL A Country Excluded from the Donor Pool						
Country	Austria	Belgium	Denmark	Finland	France	Germany	Greece
Austria	-	0	0	0	0	0	0
Belgium	0.196	-	0.044	0.241	0.197	0.244	0.261
Denmark	0	0	-	0	0	0	0
Finland	0	0.023	0	-	0	0	0
France	0	0.027	0.134	0	-	0	0
Germany	0	0	0	0	0	-	0
Greece	0.079	0.051	0	0.044	0.055	0.077	-
Ireland	0.048	0.129	0.006	0	0.020	0	0.006
Italy	0.152	0.184	0.233	0.165	0.172	0.160	0.211
Luxembourg	0.026	0.034	0.008	0.006	0.007	0.008	0.011
Netherlands	0	0	0	0	0	0	0
Portugal	0.373	0.427	0.461	0.369	0.381	0.347	0.396
Sweden	0.126	0.090	0.114	0.175	0.164	0.164	0.115
United Kingdom	0	0.036	0	0	0.003	0	0

	PANEL B Country Excluded from the Donor Pool						
Country	Ireland	Italy	Luxem.	Nether.	Portugal	Sweden	UK
Austria	0	0	0	0	0	0	0
Belgium	0.140	0.285	0.159	0.236	0.352	0.307	0.244
Denmark	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0.016	0
France	0.128	0.114	0.085	0	0.334	0.111	0
Germany	0	0	0	0	0	0	0
Greece	0.007	0.098	0.072	0.052	0.214	0.123	0.067
Ireland	-	0	0	0	0.003	0.030	0
Italy	0.178	-	0.180	0.173	0.032	0	0.172
Luxembourg	0	0	-	0	0	0	0
Netherlands	0	0	0	-	0	0.190	0
Portugal	0.407	0.289	0.353	0.374	-	0.222	0.358
Sweden	0.140	0.125	0.135	0.165	0	-	0.159
United Kingdom	0	0.090	0.0170	0	0.066	0	-

Notes: This table reports the weights assigned to each country using the synthetic control method explained in Section 3.3. In this case, we take out from the donor pool one country at the time.

Bibliography

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and student achievement in the Chicago public high schools," *Journal of labor Economics*, 2007, *25* (1), 95–135.

Abadie, Alberto, "Semiparametric instrumental variable estimation of treatment response models," *Journal of econometrics*, 2003, *113* (2), 231–263.

_ , Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 2010, *105*, 493–505.

_ and Javier Gardeazábal, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 2003, *93* (1), 113–132.

Abellán, José M., Jorge E. Martínez, Ildefonso Méndez, Jose L. Pinto, and Fernando I. Sánchez, "El Valor Monetario de Una Vida Estadística en España: Estimación en el Contexto de los Accidentes de Tráfico," Technical Report, Report Commisioned by the Spanish Directorate General of Traffic (DGT) 2009.

Abouk, Rahi and Scott Adams, "Texting bans and fatal accidents on roadways: Do they work? Or do drivers just react to announcements of bans?," *American Economic Journal: Applied Economics*, 2013, *5* (2), 179–199.

Adams, Scott and Chad Cotti, "Drunk driving after the passage of smoking bans in bars," *Journal of Public Economics*, 2008, *92* (5), 1288–1305.

_ , McKinley L Blackburn, and Chad D Cotti, "Minimum wages and alcohol-related traffic fatalities among teens," *Review of Economics and Statistics*, 2012, *94* (3), 828–840.

Aizer, Anna and Joseph J. Doyle, "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges," *The Quarterly Journal of Economics*, 2015.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2008.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 1999, *114* (2), 533–575.

_ , Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters, "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics*, 2016, *34* (2), 275–318.

Angrist, Joshua, Victor Lavy, and Analia Schlosser, "Multiple experiments for the causal link between the quantity and quality of children," *Journal of Labor Economics*, 2010, *28* (4), 773–824.

Arulampalam, Wiji, Robin A Naylor, and Jeremy Smith, "Am I missing something? The effects of absence from class on student performance," *Economics of Education Review*, 2012, *31* (4), 363–375.

Ashenfelter, Michael Greenstone Orley, "Using Mandated Speed Limits to Measure the Value of a Statistical Life," *Journal of Political Economy*, 2004, *112* (S1), S226–S267.

Ashenfelter, Orley, "Measuring the value of a statistical life: problems and prospects," *The Economic Journal*, 2006, *116* (510), C10–C23.

Aucejo, Esteban M and Teresa Foy Romano, "Assessing the effect of school days and absences on test score performance," *Economics of Education Review*, 2016, *55*, 70–87.

Baker, Michael, "Industrial actions in schools: strikes and student achievement," *Canadian Journal of Economics/Revue canadienne d'économique*, 2013, *46* (3), 1014–1036.

Behaghel, Luc, Clément de Chaisemartin, and Marc Gurgand, "Ready for boarding?: the effects of a boarding school for disadvantaged students," 2015.

Belot, Michèle and Dinand Webbink, "Do Teacher Strikes Harm Educational Attainment of Students?," *Labour*, 2010, *24* (4), 391–406.

Bettinger, Eric, Bridget Long, and Eric Taylor, "When inputs are outputs: The case of graduate student instructors," *Economics of Education Review*, 2016, *52*, 63–76.

_ **, Lindsay Fox, Susanna Loeb, and Eric Taylor**, "Changing distributions: How online college classes alter student and professor performance," Technical Report, Working Paper, Stanford University 2014.

Bobonis, Gustavo J and Frederico Finan, "Neighborhood peer effects in secondary school enrollment decisions," *The Review of Economics and Statistics*, 2009, *91* (4), 695–716.

Borjas, George, "Foreign-Born Teaching Assistants and the Academic Performance of Undergraduates," *The American Economic Review*, 2000, *90* (2), 355–359.

Bourgeon, Jean-Marc and Pierre Picard, "Point-Record Driving Licence and Road Safety: An Economic Approach," *Journal of Public Economics*, 2007, *91* (1-2), 235–258.

Braga, Michela, Marco Paccagnella, and Michele Pellizzari, "The impact of college teaching on students' academic and labor market outcomes," *Journal of Labor Economics*, 2016, *34* (3), 781–822.

Brodaty, Thibault and Marc Gurgand, "Good peers or good teachers? Evidence from a French University," *Economics of Education Review*, 2016, *54*, 62–78.

Card, David, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial & Labor Relations Review*, 1990, *43* (2), 245–257.

— **and Alan B Krueger**, "Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania," *The American Economic Review*, 1994, *84* (4), 772–793.

Carrell, James E. West Scott E., "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 2010, *118* (3), 409–432.

Carrell, Scott and James West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 2010, *118* (3), 409–432.

Carrell, Scott E and Mark L Hoekstra, "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–28.

Castillo-Manzano, José I., Mercedes Castro-Nuño, and Diego J. Pedregal, "An Econometric Analysis of the Effects of the Penalty Points System Driver's License in Spain," *Accident Analysis and Prevention*, 2010, *42*, 1310–1319.

Castleman, Benjamin and Lindsay Page, "Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates?," *Journal of Economic Behavior & Organization*, 2015, *115*, 144–160.

Chen, Jennjou and Tsui-Fang Lin, "Class Attendance and Exam Performance: A Randomized Experiment," *The Journal of Economic Education*, 2008, *39* (3), 213–227.

Chetty, Raj, John Friedman, and Jonah Rockoff, "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *The American Economic Review*, 2014, *104* (9), 2633–2679.

— **, John N Friedman, and Jonah E Rockoff**, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *The American Economic Review*, 2014, *104* (9), 2593–2632.

Cohen, Alma and Liran Einav, "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," *Review of Economics and Statistics*, 2003, *85* (4), 828–843.

Connolly, Faith and Linda S Olson, "Early Elementary Performance and Attendance in Baltimore City Schools' Pre-Kindergarten and Kindergarten.," *Baltimore Education Research Consortium*, 2012.

Cuseo, Joe, "The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students," *The Journal of Faculty Development*, 2007, *21* (1), 5–21.

Dee, Thomas, "Motorcycle Helmets and Traffic Safety," *Journal of Health Economics*, 2009, *28*, 398–412.

DePaola, Maria, Vicenzo Scoppa, and Mariatiziana Falcone, "The Deterrent Effects of the Penalty Points System for Driving Offences: a Regression Discontinuity Approach," *Empirical Economics*, 2013, *45*, 965–985.

DGT, "Anuario Estadístico General," Technical Report, Dirección General de Tráfico 2009.

_ , "Cinco Años del Permiso por Puntos," Technical Report, Dirección General de Tráfico 2011.

Dionne, Georges, Jean Pinquet, Matthieu Maurice, and Charles Vanasse, "Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data," *Review of Economics and Statistics*, 2011, *93* (1), 218–227.

Dobkin, Carlos, Ricard Gil, and Justin Marion, "Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment," *Economics of Education Review*, 2010, *29* (4), 566–575.

Duflo, Esther, Rema Hanna, and Stephen P Ryan, "Incentives work: Getting teachers to come to school," *American Economic Review*, 2012, *102* (4), 1241–78.

Durden, Garey C and Larry V Ellis, "The effects of attendance on student learning in principles of economics," *The American Economic Review*, 1995, *85* (2), 343–346.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth, "The long-run economic consequences of high-stakes examinations: evidence from transitory variation in pollution," *American Economic Journal: Applied Economics*, 2016, *8* (4), 36–65.

Evans, Leonard, "The Effectiveness of Safety Belts in Preventing Fatalities," *Accident Analysis and Prevention*, 1986, *18* (3), 229–241.

Feld, Jan, Nicolas Salamanca, and Ulf Zölitz, "Students are almost as effective as professors in university teaching," Technical Report, Unpublished Manuscript 2017.

Fleisher, Belton, Masanori Hashimoto, and Bruce Weinberg, "Foreign GTAs can be effective teachers of economics," *The Journal of Economic Education*, 2002, *33* (4), 299–325.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek, "Long-term effects of class size," *The Quarterly Journal of Economics*, 2012, *128* (1), 249–285.

González, Felipe, "Collective Action in Networks: Evidence from the Chilean Student Movement," 2016.

González, S and J Montealegre, "Ciudadanía en marcha: Educación superior y movimiento estudiantil 2011: Curso y lecciones de un conflicto," *Santiago, Chile: USACH*, 2012.

Goodman, Joshua, "Flaking out: Student absences and snow days as disruptions of instructional time," Technical Report, National Bureau of Economic Research 2014.

Gottfried, Michael A., "Evaluating the Relationship Between Student Attendance and Achievement in Urban Elementary and Middle Schools: An Instrumental Variables Approach," *American Educational Research Journal*, 2010, *47* (2), 434–465.

Green, Colin, John Heywood, and Maria Navarro, "Did Liberalising Bar Hours Decrease Traffic Accidents?," *Journal of Health Economics*, 2014, *35*, 189–198.

Herrmann, Mariesa A and Jonah E Rockoff, "Worker absence and productivity: Evidence from teaching," *Journal of Labor Economics*, 2012, *30* (4), 749–782.

Hoffmann, Florian and Philip Oreopoulos, "Professor qualities and student achievement," *The Review of Economics and Statistics*, 2009, *91* (1), 83–92.

Hoxby, Caroline, "How teachers' unions affect education production," *The Quarterly Journal of Economics*, 1996, pp. 671–718.

Hsieh, Chang-Tai and Miguel Urquiola, "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of public Economics*, 2006, *90* (8), 1477–1503.

IEA, "TIMSS and TIMSS advanced 2015 international results," Lynch School of Education, Boston College. 2015.

Jacob, Brian A, Lars Lefgren, and David P Sims, "The persistence of teacher-induced learning," *Journal of Human resources*, 2010, *45* (4), 915–943.

Kane, Thomas and Douglas Staiger, "Estimating teacher impacts on student achievement: An experimental evaluation," Technical Report, National Bureau of Economic Research 2008.

Kokkelenberg, Edward C, Michael Dillon, and Sean M Christy, "The effects of class size on student grades at a public university," *Economics of Education Review*, 2008, *27* (2), 221–233.

Krueger, Alan B, "Economic considerations and class size," *The Economic Journal*, 2003, *113* (485), F34–F63.

Lavy, Victor, "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *The American Economic Review*, 2009, *99* (5), 1979–2011.

_ , "Expanding school resources and increasing time on task: Effects of a policy experiment in Israel on student academic achievement and behavior," Technical Report, National Bureau of Economic Research 2012.

_ , "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries," *The Economic Journal*, 2015, *125* (588), F397–F424.

_ , Avraham Ebenstein, and Sefi Roth, "The Long Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution," *American Economic Journal: Applied Economics, forthcoming*, 2015.

Lazear, Edward P., "Educational Production," *The Quarterly Journal of Economics*, 2001, *116* (3), 777–803.

Lovenheim, Michael and Joel Slemrod, "The Fatal Toll of Driving to Drink: The Effect of Minimum Legal Drinking Age Evasion on Traffic Fatalities," *Journal of Health Economics*, 2010, *29*, 62–77.

Lusher, Lester, Doug Campbell, and Scott Carrell, "TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates," Technical Report, National Bureau of Economic Research 2015.

Martins, Pedro and Ian Walker, "Student achievement and university classes: Effects of attendance, size, peers, and teachers," *IZA Discussion Paper No. 2490*, 2006.

Maurin, Eric and Sandra McNally, "Vive la Révolution! Long-Term Educational Returns of 1968 to the Angry Students," *Journal of Labor Economics*, 2008, *26* (1), 1–33.

McCluskey, Cynthia Perez, Timothy S Bynum, and Justin W Patchin, "Reducing chronic absenteeism: An assessment of an early truancy initiative," *Crime & Delinquency*, 2004, *50* (2), 214–234.

MINEDUC, "Variación de matrícula y tasas de permanencia por sector," Ministerio de Educación, Gobierno de Chile. 2015.

Neilson, Christopher, "Targeted vouchers, competition among schools, and the academic achievement of poor students," *Documento de trabajo). Yale University. Recuperado de http://economics. sas. upenn. edu/system/files/event_papers/Neilson_ 2013_ JMP_ current. pdf*, 2013.

Novoa, Ana M., Katherine Pérez, Elena Santamariña-Rubio, Marc Marí-Dell'Olmo, Josep Ferrando, Rosana Peiró, Aurelio Tobías, Pilar Zori, and Carme Borrell, "Impact of the Penalty Points System on Road Traffic Injuries in Spain: A Time–Series Study," *American Journal of Public Health*, 2010, *100* (11), 2220–2227.

OECD, "The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence," Organisation for Economic Co-operation and Development, France. 2015.

_ , "PISA: results in focus," Organisation for Economic Co-operation and Development, France. 2016.

Peltzman, Sam, "The Effects of Automobile Saftey Regulation," *Journal of Political Economy*, 1975, *83* (41), 677–725.

Pischke, Jörn-Steffen, "The impact of length of the school year on student performance and earnings: Evidence from the German short school years," *The Economic Journal*, 2007, *117* (523), 1216–1242.

Pulido, José, Pablo Lardelli, Luis de la Fuente, Victor M. Flores, Fernando Vallejo, and Enrique Regidor, "Impact of the Demerit Point System on Road Traffic Accident Mortality in Spain," *Journal of Epidemiology Community Health*, 2010, *64*, 274–276.

Reyes, Loreto, Jorge Rodríguez, and Sergio S Urzúa, "Heterogeneous economic returns to postsecondary degrees: Evidence from chile," Technical Report, National Bureau of Economic Research 2013.

Rivkin, Steven G, Eric A Hanushek, and John F Kain, "Teachers, schools, and academic achievement," *Econometrica*, 2005, *73* (2), 417–458.

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

Romer, David, "Do students go to class? Should they?," *The Journal of Economic Perspectives*, 1993, *7* (3), 167–174.

Romero, Maria José and Young-Sun Lee, "A national portrait of chronic absenteeism in the early grades," 2007.

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

_ , "Teacher quality in educational production: Tracking, decay, and student achievement," *The Quarterly Journal of Economics*, 2010, *125* (1), 175–214.

_ , "Teacher quality policy when supply matters," *The American Economic Review*, 2014, *105* (1), 100–130.

Schanzenbach, Diane W, "Does Class Size Matter?," Technical Report, National Education Policy Center, School of Education, University of Colorado, Boulder 2014.

Simonsen, Elizabeth, *Mala Educación: Historia de la revolución escolar*, Random House Mondadori, 2012.

Solis, Alex, "Credit access and college enrollment," Technical Report, Working Paper, Department of Economics, Uppsala University 2013.

Stanca, Luca, "The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics," *The Journal of Economic Education*, 2006, *37* (3), 251–266.

UNICEF, "La Voz del Movimiento Estudiantial 2011. Educación Pública, Gratuíta y de Calidad.," Fomdo de las Naciones Unidas para la Infancia. 2014.

US Department of Education, "2013-2014 Civil rights data collection: a first look," Department of Education, The United States of America. 2016.

Veloso, Nicolas A Grau, "Two essays on the economics of education," 2014.

Viscusi, Kip, "The value of risks to life and health," *Journal of economic literature*, 1993, *31* (4), 1912–1946.

— **and Joseph Aldy**, "The value of a statistical life: a critical review of market estimates throughout the world," *Journal of risk and uncertainty*, 2003, *27* (1), 5–76.

Vlieger, Pieter De, Brian Jacob, and Kevin Stange, "Measuring Instructor Effectiveness in Higher Education," in "Productivity in Higher Education," University of Chicago Press, 2017.

Waldinger, Fabian, "Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany," *The Review of Economic Studies*, 2012, *79* (2), 838–861.

WHO, "Global Status Report on Road Safety," Technical Report, World Health Organization 2013.