The London School of Economics and Political Science

Essays in Economics of Education

Marta De Philippis

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy. London, February 2016

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of 34,123 words.

Statement of inclusion of previous work

I confirm that Chapter 2 is a heavily revised version of the paper I submitted at the end of my MRes in 2012.

Marta De Philippis

Statement of conjoint work

I confirm that Chapter 3 was jointly co-authored with Federico Rossi, PhD student at the London School of Economics. Each of us contributed equally to the creation, development and writing of the chapter.

Marta De Philippis *Federico Rossi*

Acknowledgements

First and foremost, I am deeply indebted to my supervisor Prof. Steve Pischke for his valuable guidance and thoughtful insights, but also for his encouragement and understanding during these years. He has taught me, both consciously and unconsciously, how good economics is done and to never stop questioning my results.

I am also extremely grateful to Esteban Aucejo, for the many stimulating discussions we had together. The enthusiasm he has for his research was contagious and motivational for me, even during tough times in the PhD pursuit.

Moreover I want to thank my co-author Federico Rossi for his work and energy. I always say that I chose the best co-author I could have had.

I also wish to thank all members of the LSE labour faculty, of the CEP labour and education group and of the labour division at the Bank of Italy, for fruitful advices and suggestions on my research work.

I gratefully acknowledge Unicredit and Universities Foundation and the ESRC for their generous financial support over the course of these years.

LSE PhD students and CEP researchers have contributed immensely to my personal and professional time at the LSE. The group has been a source of friendships as well as good advice and collaboration. A special mention goes to Giulia, Georg, Stephan and Marco for their constant help and support.

I also thank my "family" in London: Cecilia, Barbara, Damiano, Ferdinando, Enrico and Luca. We went through hard times together, we cheered us on, and we celebrated each accomplishment of these last 5 years. They will always have a special place in my heart.

I thank Ludovico for his support and his unique ability to make me laugh and to help me putting things in perspective. And because he always thought I would make it.

Last but not least, I would like to thank my family for their love, support and infinite patience during these years. They contributed to this thesis in a million of different ways but I will never forget their words while driving me at the airport after the first term at the LSE. Thank you.

Introduction

This thesis studies aspects related to the role of schools characteristics and their governance on students' learning outcomes. The thesis contains three chapters. The first chapter explores the effect of exposing students to more science in high school on their enrolment and persistence in STEM majors at university. It exploits the different timing in the implementation of a reform that induced high schools in the UK to offer more science to high ability 14 year-old children. The findings show that a stronger science curriculum at high school increases the probability of enrolling and of graduating in a STEM major at university. Moreover, the effect masks substantial gender heterogeneity. It is indeed mostly concentrated on boys. Girls tend to choose more scientific subjects, but still the most female-dominated ones: they choose medicine, not engineering.

The second chapter of this thesis analyses the effects of providing strong research incentives to university professors on the way they allocate effort between teaching and research and on the way they select into different types of universities. I find evidence that teaching and research efforts are substitute in the professors' cost function: the impact of research incentives is positive on research activity and negative on teaching performance. Effects are stronger for young faculty members, who are exposed not only to monetary incentives but also to career concerns. Moreover, I find that less skilled researchers tend to leave the university under stronger research incentives. Since I estimate that teaching and research skills are positively correlated, this implies that also bad teachers tend to leave the university. The overall impact of stronger research incentives on the university teaching quality is therefore ambiguous: the negative effect on teaching performance for incumbent professors is compensated by the positive sorting effect, given by changes in the composition of teachers.

The third chapter explores where do the large cross-country differences in students' performances in international standardized tests come from. This chapter argues that, while most of the debate concentrates on country differences in the school systems, differences in cultural environments and parental inputs are instead of great importance. I show indeed that the school performance of second generation immigrants is closely related to the average one of native students who still study in their parents' countries of origin. This holds true even after accounting for different family background characteristics, different schools attended and different patterns of selection into immigration. This pattern questions whether PISA scores should be interpreted only as a quality measure for a country's educational system. They actually contain an important intergenerational and cultural component. Parental inputs are found indeed to explain a large part of the cross country variation in school performance, for instance they account for more than one third of the gap between East Asia and other regions.

Contents

Chapter 1

STEM Graduates and High School Curriculum: Does Early Exposure to Science Matter?

1.1 Introduction

In the new heavily globalized and innovation driven economy, increasing the number of Science, Technology, Engineering and Math (STEM)[1] graduates at university is found to generate high social returns in terms of long-term productivity, growth and competitiveness (Winters, 2014; Peri et al., 2013; Moretti, 2012; Atkinson and Mayo, 2010; Jones, 2002). Moreover, graduating in a STEM major also represents a very profitable private investment for college graduates themselves. First, lifetime earnings of graduates in STEM majors are extremely high (Kirkeboen et al., 2016; Hastings et al., 2013; Pavan and Kinsler, 2015; Rendall and Rendall, 2014; Koedel and Tyhurst, 2010): Altonji et al. (2012) show that nowadays intra-educational income differences are comparable to inter-educational differences. In the US in 2009 the wage gap between the average electrical engineer and someone majored in general education was almost identical to the wage gap between the average college graduate and the average high school graduate. Second, the differences in earnings among university majors have increased substantially over time: some scholars (Rendall and Rendall, 2014) claim that a large part of the increased income inequality in the US is driven by increasing

[1] Throughout the paper I define as "STEM" the following majors: Physical science, Mathematical and Computer science and Engineering.

returns of scientific and mathematical university majors. Third, graduates in STEM fields earn more independently of the quality of the institution they attend (Kirkeboen et al., 2016; Arcidiacono et al., forthcoming): engineering is considered a better investment than Harvard (James et al., 1989). Finally, also non-monetary returns are high in STEM occupations: Goldin (2014) classifies occupations based on their degree of temporal flexibility, i.e. how important it is to stay long or particular hours in the office, and STEM occupations are ranked among the first.

However, despite the high social and private benefits obtained from graduating in STEM majors, the general consensus among policy-makers is that the current supply of STEM skills is insufficient and, when combined with forecast growth in demand for STEM skills, it presents a potentially significant constraint on future economic growth (UK HM Treasury and BIS, 2010; The President's Council of Advisor on Science and Technology, 2012; European Commission, 2010).[2] The governments of many countries invested a very large amount of funds to induce more graduates in STEM (Atkinson and Mayo, 2010). The US federal government for instance is considering actions with the objective of increasing STEM graduates by 34% annually (The President's Council of Advisor on Science and Technology, 2012). Still, the graduation rate or even the degree of interest for students in graduating in these majors has remained pretty stable since the '80s (Altonji et al., 2012) and, while the literature on choices of educational levels is very wide and consolidated (starting from the seminal work by Mincer (1974)), there is relatively little work on choices of fields of study.

This paper evaluates how much of the lack in STEM graduates can be attributed to high schools, and in particular to the curriculum they offer. Ellison and Swanson (2012) show indeed that there is a large heterogeneity in high schools effectiveness in developing talents in math and science, which is not explained by differences in schools composition. This paper investigates the role of high school curriculum and it addresses three questions in particular. First, does being exposed to more science courses in high school increase by itself the supply of STEM graduates? Second, who is induced to take more science classes, when exposed to the option of taking them at age 14? Third, this paper evaluates whether more exposure to science at age 14 works for everybody, or whether the effect is concentrated on some segments of the population. This is a relevant question to distinguish whether is it more efficient to force all students to take more science courses in high school or to target the offer

[2] Overall, STEM employment grew three times more than non-STEM employment over the last twelve years, and it is expected to grow twice as fast by 2018. According to a report by the Information Technology and Innovation Foundation (2010), the number of STEM graduates in the US will have to increase by 20-30% by 2016 to meet the projected growth of the economy.

to some subgroups of the population only, those most likely to benefit from it. Some papers (Dougherty et al., 2015) find that universal coverage of stronger mathematics courses may not be overall beneficial. Moreover, by evaluating the heterogeneity of the effect with respect to gender, this chapter investigates whether increasing students' preparation in science has an effect on shrinking the gender gap in enrollment in STEM majors.

The identification of the effect of studying more science in high school is difficult because of what I call a 'nested selection' problem: of students into different schools -based on the curriculum they offer- and of students into different courses within the school they chose. To my knowledge I am the first to fully address and test both sources of endogeneity to answer this type of question. I eliminate the selection in different courses within the same school by collapsing the analysis at the school level (in the spirit of Altonji (1995)). I address the other layer of selection, i.e. the selection of students into different schools, by exploiting exogenous variation in the timing of the introduction of an advanced science course in English high schools. The UK government introduced in 2004[3] an entitlement to study advanced science for high ability students at age 14, with the explicit aim of fostering enrollment in post-secondary science education. This resulted in a strong increase in the number of schools offering advanced science: from 20% in 2002 to 80% in 2011. As a consequence, the share of students taking advanced science increased from 4% in 2002 to 20% in 2011 and the increase was almost entirely concentrated on high ability students[4] (see Figure 1). Thanks to a novel dataset, obtained combining different administrative sources from the UK, I propose two alternative identification strategies that approach this type of selection problem from two complement perspectives and use two different sources of variation. The first uses within school over time variation in the type of courses offered. It exploits the three year time lag between the moment when students choose their high school (age 11) and the moment when they choose their field courses (age 14). It evaluates the effect on students unexpectedly exposed to the offer of the advanced science course, because their schools started to offer the advanced science course only once they chose the school.[5]

The second identification strategy tests the robustness of my results by using variation in whether the schools are offering advanced science that were in place even before the students started to attend their schools. I exploit the fact that schools in the UK, when

[3]After the publication on a ten-year investment framework for science and innovation (UK Government, 2004)

[4]I define high ability students as those who were in the top 30 percentile of the primary school grades distribution. The increase for these students was of 35 percentage points, from 15% to about 50%.

[5]A similar idea, with only one year lag, has been used by Joensen and Nielsen (2009, 2015), to evaluate a different treatment. However in their context schools self-selected in the program and, due to the lack of data, it was impossible to check that selection was not driven by cohort-specific demand-driven shocks or by the fact that families anticipated the change in curriculum (since there is 1 year lag only).

oversubscribed, select students based to home-to-school distance and schools catchment areas vary (unpredictably) over time. My second instrument therefore compares students living in the same neighbourhood, who are more or less likely to enrol in schools offering advanced science, because of exogenous changes in schools' catchment areas over time.

My empirical findings may be summarized by three broad conclusions. First, I find that taking advanced science at age 14 increases the probability of choosing science at age 16 by 5 percentage points and of enrolling in STEM majors by about 2 percentage points. Moreover, offering more science courses at high school not only induces more students to enrol in STEM majors but also increases the likelihood that they graduate in these majors. This is extremely important, given the large problem with persistence in this kind of majors (Arcidiacono et al., forthcoming; Stinebrickner and Stinebrickner, 2014). Second, I find that previous test scores in science are the main determinant behind the choice of taking advanced science, when students are exposed to this option. Boys and girls at this stage behave similarly and select into advanced science in the same proportion. Still, third, I find that the final effects are very heterogeneous and are mainly concentrated on boys and on middle-high ability students. The gender gap in STEM major enrolment therefore increases as a consequence of this policy; not because less girls take advanced science at age 14, but because girls, when exposed to more science in high school, are induced to take, yes, more challenging subjects[6], but still the most female-dominated ones: they choose medicine, not engineering.

Taken together, my findings can inform ongoing debates over government intervention to address apparent mismatches and market frictions in the supply and demand for post-secondary fields of study. My results suggest that to reinvigorate STEM education, and high skilled STEM education in particular, governments should consider a policy aimed at offering more science courses to high ability students during high schools. I estimate that the policy contributed to one third of the increase the share of STEM graduates in the UK between 2005 and 2010.

This paper speaks to the literature aimed at explaining choices of university majors. Most of what we know so far comes from surveys or informational experiments. The evidence is mixed. The most common explanations look at the role of expected earnings; of competencies and preparation; of self-confidence; of preferences and of innate ability (Arcidiacono et al., 2012; Arcidiacono, 2004; Beffy et al., 2012; Stinebrickner and Stinebrickner, 2014; M and Zafar, 2014). However preferences and ability are usually considered to be fixed over time,

[6] I define challenging the subjects usually taken by high ability students

and it is therefore difficult for policy-makers to shape them; returns to STEM majors are already very high, as stated before, and the elasticity of major choice to expected earnings is found to be very low (Beffy et al., 2012). Moreover, Stinebrickner and Stinebrickner (2014) show that students start university being over-confident not under-confident about their scientific ability. There is, instead, large scope for policies intervening on students' preparation and on primary and secondary schools quality. Many scholars (Cameron and Heckma, 2001; Moretti, 2012), indeed, attribute the lack of STEM graduates to the low quality of the US school system. While some studies look at the effects of school inputs (usually at the university level), like peers (De Giorgi et al., 2010; Anelli and Peri, 2015), teachers (Scott E. Carrell and West, 2010) and university coursework (Fricke et al., 2015), there is surprisingly little quantitative work on the effects of high school curriculum (Altonji et al., 2012). Still, not only every single government has to make at some point the decision about how to design its country high school curriculum and how to implement it optimally in order to reduce possible mismatches between demand and supply of skills, but also, differently from other policies like changes in peers, this is not a zero sum choice: everybody may potentially benefit from an optimally designed curriculum.

My paper improves on the existing literature in several ways.[7]
First, I address both layers of selection of students into courses mentioned before. Most studies (Altonji, 1995; Levine and Zimmerman, 1995; Betts and Rose, 2004; Joensen and Nielsen, 2009, 2015) use across school variation in the type of curriculum offered and do not fully address possible selection of students into schools, based on the curriculum they offer. Since family background and individual motivation are important determinants of both choice of majors and of high schools, the bias in estimates that do not take into account selection into schools could be important and could lead to overestimate the effects. I show that, even in my context when the variation in curriculum is induced by a policy, adding school-level controls is not enough to eliminate selection bias: when checking the identifying assumptions, the inclusion of school fixed effects and the presence of an instrument turn out to be crucial to correctly identify the effect of interest.
Second, I am able to identify the effect of offering more (natural) science courses only. Usually (Altonji, 1995; Joensen and Nielsen, 2009, 2015; Gorlitz and Gravert, 2015; Jia, 2014), changes in high school curricula imply a restructuring of many different courses and it is difficult to isolate the effect of one single subject. The policy I consider strengthens instead the natural science curriculum only, without intervening on other subjects. Still, my treatment has multiple components: taking advanced science also implies a change in

[7] I mention here both papers that look at the effect on earnings and on majors, even if most of the literature looks at earnings without focussing on the effect on choices of majors.

classroom heterogeneity and composition.[8] I disentangle the curriculum effect from the peer channel, using an instrument for peers that exploits within-school over-time variation in the ability of predicted peers, depending on whether the school offers advanced science or not. I find that the effect of the additional science course persists even after controlling for changes in peers' characteristics. This is key to identify the exact origin of the effect and therefore to allow policy-makers to reproduce the policy in other contexts.
Third, I am able to look at the effect for extremely high ability students with potentially very high probability of succeeding in STEM majors. These are the students of highest interest for policy-makers because they are more likely to actually enrol in STEM majors and to make important contributions to scientific and technological fields. Most of the existing empirical works (Goodman, 2012; Cortes et al., 2015) analyze policies that affect almost entirely low ability students, not likely to enroll to university at all, or students for whom taking science is rather costly (Joensen and Nielsen, 2009, 2015).[9]

The remainder of this paper is organized as follows. In Section 1.2, I describe the data, the English school system and the reform of the advanced science program in UK high schools. Section 1.3 provides an overview of the main identification strategies. Section 1.4 presents the estimated impact of advanced science on post-16 educational outcomes and it checks the identifying assumptions and the robustness of the results. Section 1.5 inspects the mechanisms behind the estimates and, finally, Section 1.6 concludes.

1.2 Data and institutional setting

1.2.1 The English school system

Compulsory education in England is organized in four Key Stages (KS). At the end of each stage students are evaluated in standardized national exams. Figure 2 shows a timeline of the UK educational system. Pupils enter school at age 4, the Foundation Stage, then they move to Key Stage 1 (KS1), spanning ages 5 and 6, and Key Stage 2 (KS2, from age 7 to age

[8]because the advanced science course provides the possibility of taking a course exclusively attended only by other very high ability students.

[9]These studies exploit for instance changes in minimum math requirements across US states over time or compare students just below or just above the threshold for attending remedial classes in math and find modest effects on earnings, concentrated on low-SES students. In my setting, instead, compliers include also extremely high ability students, within the same school.

11).[10] At the end of KS2 children leave primary school and go to secondary school, where they progress to Key Stage 3 (KS3, age 12-14) and Key Stage 4 (KS4, age 15-16). At KS4 students start choosing some subjects.[11] In particular, out of usually between 10 and 12 qualifications, students typically choose between 4 and 6 subjects.[12] At age 16 compulsory education ends and students may continue their secondary studies for a further two years. This phase is called Key Stage 5 (age 17-18) and may take place in the same secondary school (about 60% of the schools also offer KS5 courses) or in a different school. Again, students have many different options: they can choose more vocational or more academic-oriented type of qualifications (the so-called A levels). Students usually take three A level or equivalent qualifications[13], and are free to choose any subject.[14] Finally, higher education usually begins at age 19 with a three-year bachelor's degree. Admission to university is usually based on which subjects were chosen at KS5 and on the grades achieved.

1.2.2 Science in high school

While science is a core component of the National Curriculum at KS4, there are several different ways to fulfill the requirement. All students are required to study the basics elements of all three natural sciences (physics, chemistry and biology) and should at least take the so-called 'single science' or core science course (which is worth 1 KS4 qualification). They can, moreover, choose to take the 'double science' course (worth 2 qualifications) which leads to more knowledge in all the three subjects or the 'triple science' course (which is called advanced science and is equivalent to take one full qualification in each of the three natural science subjects). Finally students can also choose more vocational science qualifications. Taking triple science implies both longer instruction time and the study of more complex science topics.[15] Double science and, more recently, triple science provide the standard routes into the fulfillment of KS4 requirements.

[10] KS1 corresponds to grade 1 and 2 in the US school system, KS2 to grades 3,4 and 5.

[11] A number of different qualification types are available to young people at KS4, varying in their level of difficulty. These include: GCSE (the most common qualification in the UK and the most academic oriented), and other more vocational qualifications. I will only consider GCSE qualifications or GCSE equivalent qualifications.

[12] The six compulsory subjects are: English, math, (single) science, information and communication, physical education and citizenship. Students in general take overall between 10 and 12 qualifications.

[13] 50% of students takes between 3 and 3.5 A level equivalent qualifications.

[14] Many university, when selecting students, look not only at grades in A level but also at the subjects chosen, however none of them takes into consideration the type of KS4 subjects chose (see Section 1.4.4).

[15] In this case students study more difficult topics such as electric current, transformers, some medical application, more quantitative topics in chemistry etc.

In 2004 the UK Government published a ten-year investment framework for science and innovation (UK Government, 2004). The framework set out the Governments ambition for UK science and innovation over the next decade and emphasized in particular the need for more graduates in science. Taking triple science was considered extremely important, because "it gives students the necessary preparation and confidence to go on and study science" (Confederation of British Industry). An entitlement to study triple science for students achieving at least level 6 or above at KS3 science (the students on the top 40% of the grade distribution) was established.[16] The result was a very large increase in the number of schools offering triple science. While in 2002 less than 20% of schools offered triple science, by 2011 the share became more than 80% (see Figure 1). Between 2002 and 2011 the share of students choosing triple science increased from 4% to 20% and the increase was mostly concentrated among high ability students[17] (for whom the share increased from 15% to 50%).

Since I will exploit within school over time variation in courses offered, it is important to explore what drives this variation.There are several, mainly supply driven, reasons why the exact timing of the introduction of the triple science option differs by schools. First, the lack of specialized teachers. 50% of science and math students in English high schools are not taught by teachers specialized in the subject. For teachers teaching outside their expertise, triple science is particularly demanding and they need more time to get familiar with the material. Second, the school size: for small schools it is difficult to offer a large amount of subject options. With the ten-year investment framework, the government encouraged new collaborative arrangements with other schools (to jointly provide triple science). However, setting these agreements up takes time and many schools need the support of their Local Education Authority (LEA) and the exact timing of the conclusion of these agreements is uncertain. Finally, support and pressure on schools to fulfill the entitlement to triple science was provided at the LEA level.[18] Some LEAs are not as positive and supportive as others regarding the introduction of triple science. Figure 3 shows that the increase in the share of schools offering triple science was very heterogeneous across different LEAs.

[16]In particular the government stated that "all pupils achieving at least level 6 [Level 6 or above is equivalent to the top 30% of students] at KS3 should be entitled to study triple science at KS4, for example through collaborative arrangements with other schools.

[17]Based on their primary school grade in math, science and English.

[18]LEAs organize courses both on how to organize the time schedule to fit the new curriculum and on the new material covered and encourage school-to-school learning. There is large heterogeneity on how actively different LEAs promoted and pushed the introduction of the Triple Science option in schools. In total there are 152 local authorities in England.

1.2.3 Data

I use administrative data on all students in maintained schools in the UK, from primary school till the end of secondary school. Moreover, I link these data to the universe of all UK university students. In this way, my final dataset follows students from primary school till the end of their university career.

I obtain information on students demographic characteristics from The Pupil Level Annual School Census (PLASC) that collects information on students' gender, ethnicity, Free School Meal Eligibility (FSM), Special Education Needs (SEN), language group as well as their postcodes.

I obtain information on students' attainments in all their Key Stages exams (from KS1 till KS5) from the National Pupil Database (NPD). The NPD contains moreover information on every single subject chosen (and the corresponding grade) in KS4 and KS5. Finally, NPD provides a very rich set of data on school characteristics (peer groups, types of school, teachers' hirings, schools location etc.).

From the NPD dataset I also obtain the information on which courses are offered by each school. In particular, I follow the official methodology used by the English Department of Education and I infer that a school offers a course if at least one pupil at the school took an assessment in that specific course and year. My results are robust to different definitions (at least 5 pupils, at least 5% of the students, for at least two consecutive years etc.) and all different definitions are extremely highly correlated.

I then link the NPD to the universe of UK university students, the Higher Education Statistical Agency (HESA) dataset. The HESA dataset provides information on whether pupils progress to university, on their major, on the institution they attend and on whether they graduate and in which major. I combine these two data sources to create a dataset following the entire population of five cohorts of English school children. My sample includes pupils who took KS4 examinations (at age 16) between the academic years 2004/2005 and 2009/2010. After 2010, there would be no information on university outcomes, because I only have data on university results till 2013. Before 2005, there is no information on whether the school was offering triple science when the student applied to the school, because the data collection starts in 2002 and there are three years of lag. Using information on the high school attended by each individual, I match the individual record with school level data on whether the school was offering triple science when the student applied and three years later, when she had to choose her KS4 subjects.

Finally, I impose a set of standard restrictions on the data. First, I exclude special schools, hospital schools, schools where there is a three tier system instead of a two tier system. Second, I only use students who can be tracked from KS2 to KS4.[19] This leaves me with approximately 530,000 students per cohort.

The data I use are a major improvement over previous studies. While the very detailed nature of the information needed on subject choices gives particularly large scope for measurement error problems in survey data, the students' administrative dataset usually available in other countries do not usually contain all the elements necessary for this analysis. For instance, most datasets do not have information on university outcomes and the few administrative datasets that include post high school outcomes as well, refer to small countries such as Scandinavian countries, relatively homogeneous in terms of students' background and sometimes do not include information on previous test scores (Joensen and Nielsen, 2009, 2015). The large amount of observations and the heterogeneity in the students' background available in the UK dataset, provide me with enough power to accurately run my analysis and to study the heterogeneity of the effect on subgroups of the population.

1.3 Empirical strategy

1.3.1 A 'nested' selection problem

The main identification challenge when studying the effects of high school courses on post high school outcomes, is to correct for selection bias.

To fix ideas, consider the case in which students choose between taking more science in high school ($D = 1$) or not ($D = 0$). The observed choice of university major (Y) can be linked to potential majors (Y_j where $j = 1, 0$) and the type of science in high school (D) as:

$$Y = Y_0 + D(Y_1 - Y_0) \tag{1}$$

The OLS estimates of the effect of choosing more science in high school, can be written as follows:

$$E(Y|D = 1) - E(Y|D = 0) = E(Y_1|D = 1) - E(Y_0|D = 0) \tag{2}$$

[19] I checked whether this selection generates any bias (i.e. is correlated with the instrument) and this is not the case. The results are available upon request.

The main challenge is that students selecting into certain high school courses would have different potential outcomes in any case, meaning that a simple OLS does not provide the right counterfactual ($E(Y_0|D=0) \neq E(Y_0|D=1)$). In practice the bias is generated by what I call a 'nested selection' problem, because there are two layers of selection: selection of students into schools offering triple science and selection of students into triple science, for a given school.
Let's call S a dummy equal to 1 if the school attended by student i offers triple science and 0 otherwise. Then, the OLS estimates can be written as follows:

$$\begin{aligned} E(Y|D=1)-E(Y|D=0) = {} & \underbrace{E(Y_1 - Y_0|D=1, S=1)}_{\text{ATT}} + \\ & P(S=1|D=0)\,\underbrace{[E(Y_0|D=1, S=1) - E(Y_0|D=0, S=1)]}_{\text{selection into courses}} + \\ & P(S=0|D=0)\,\underbrace{[E(Y_0|D=1, S=1) - E(Y_0|D=0, S=0)]}_{\text{selection into schools+courses}} \end{aligned} \tag{3}$$

In particular, if I parametrize potential outcomes and I introduce covariates X_{ist} such that

$$Y^j_{ist} = \beta_1 D_{ist} + \beta_2 X_{ist} + \delta_s + \delta_t + u^j_{ist} \tag{4}$$

where D_{ist} is the usual dummy equal to 1 if student i in high school s, in cohort t takes triple science and 0 otherwise; $j = 1, 0$ refers to the scenarios where D_{ist} is equal to 1 or 0 respectively; X_{ist} are school and student controls; δ_s are school fixed effects and δ_t are year fixed effects. Finally, u^j_{ist} is the error term when $j = 1$ or $j = 0$. Combining equation 1 and equation 4, I get:

$$Y_{ist} = \beta_1 D_{ist} + \beta_2 X_{ist} + \delta_s + \delta_t + \underbrace{u^0_{ist} + D_{ist}(u^1_{ist} - u^0_{ist})}_{\epsilon_{ist}} \tag{5}$$

where β_1 is the effect of more science in high school on subsequent subject choices.
Estimating equation 5 by OLS is likely to lead to inconsistent and probably upward biased estimates of β_1, because of selection bias. Adding controls to equation 5 may not be enough: students may select into schools or into subjects according to unobservable characteristics, like family expectations and individual motivation that are important determinants of both choices of majors and choices of high schools and high school courses.
I address the nested selection problem by tackling the first and the second layer of selection in two different ways. Selection of students into courses within the same schools (second

layer) is addressed by collapsing the analysis at the school level, since I use an instrument that varies only at the school-cohort level. The error term becomes $\bar{\epsilon}_{st} = \overline{u}^0_{st} + D(\overline{u}^1_{st} - \overline{u}^0_{st})$ and varies only at the school-year level. Most papers (in the spirit of Altonji (1995)) use school average curriculum as instrument and basically adopt this approach and therefore address this type of selection only.

This leaves space, however, to endogeneity due to selection of students into schools offering different curricula (first layer). Differently from most of the existing literature, I also address this layer of selection and I do it in two different ways, that exploit two different types of variation.

1.3.2 First instrument

My first identification strategy uses as instrument for D_{ist} a dummy equal to one if student i in school s and cohort t was unexpectedly exposed to the triple science option. I rely on the time span between the time when students choose high schools (age 11) and the time when they choose their optional subjects (age 14). When students choose a school, they choose the school that maximizes their expected utility, conditional on their information set at age 11. However, by the time they have to choose subjects, at age 14, some schools may have started to offer triple science as induced by the policy. I will therefore compare two types of students, a priori identical because they all selected schools not offering triple science at age 11. My treated group is composed by students whose schools unexpectedly start to offer triple science by the time they turn 14 (three years after they enrol). My control group are students whose schools does not offer triple science when they have to choose subjects at age 14.

My first stage equation is:

$$D_{ist} = \gamma_1 z_{st} + \gamma_2 X_{ist} + \zeta_s + \zeta_t + v_{istj} \tag{6}$$

where z_{st} is a dummy equal to 1 if school s was not offering triple science when students from cohort t applied to secondary schools and has started to offer triple science by the time they choose their KS4 subjects at age 14; X_{ist} are school and individual controls and ζ_s and ζ_t are school and cohort fixed effects. I only include schools not offering triple science when students applied.[20]

[20]The results are robust if I also include schools already offering triple science when students applied and

This strategy mainly relies on two assumptions.
First, the assumption that the information set of both students in the treatment and in the control group at age 11, when they choose their school, does not include the information on whether the school is going to offer triple science in the next three years. This is very likely, given the large time lapse and uncertainty on when exactly teachers/classrooms and time schedules would be ready. Moreover, students are not totally free to choose the school they want: there are exogenous geographical constraints in choosing schools in the UK, especially if schools are oversubscribed. Finally, in Section 1.4.4, I show that students enrolling in schools offering triple science are observationally identical to students enrolled in schools not offering triple science: there is no sign of strategic selection of schools based on whether they offer the advanced science course, even if the information is available to parents and students at age 11.
Second, the assumption that schools' decisions on when exactly to start offering triple science are related to supply-driven rather than demand-driven factors: schools must decide when to start offering triple science not based on the quality of the current cohort attending the school. In Section 1.2.2 I described some supply driven reasons why schools may delay the introduction of triple science. In Section 1.4.4 I show that the timing of the introduction of the triple science option is not correlated with (observable) characteristics of current students in the school. Finally, I check that school s, before starting to offer triple science, was on the same trend of all other schools.

1.3.3 Second instrument

Still, even if there is no evidence that schools decide when to offer triple science depending on observable characteristics of their current cohort, it may still be that the unobservable characteristics matter. This is impossible to test. My second strategy however is not subject to this last concern because it exploits variation in available courses that existed even before the current students started to attend their high schools. This excludes the possibility that the choice of offering triple science depends on specific shocks to the particular cohort in the school.
This instrument compares students living in the same neighbourhood but who are more or less likely to enrol in schools offering triple science, because of exogenous changes in schools' catchment areas.
I exploit the fact that schools in the UK, when oversubscribed, have to prioritize based on

I control for the year they started to offer it.

geographical distance.[21] Therefore, in each year there will be a maximum distance between the school and the students' addresses above which students will not be accepted. I build my instrument in two steps: first, I compute the school catchment areas for each year, the area delimited by the circle whose centre is the school and ray is the maximum observed home-to-school distance,[22] and I define the set of 'reachable' schools for each student. Second, I compute the share of 'reachable' schools that offered triple science when student i applied. Figure 4 shows how the instrument is constructed. Address 1 refers to the lower level output area (LLOA)[23] where student i used to live at age 10. Around i's house there are three schools with different catchment areas, whose ray is indicated by the black dashed line. The instrument used in this section of the analysis counts how many schools, out of the two schools reachable by students i in year t, offered triple science when i applied to high school (in this case the instrument in year t would be 0.5). The instrument varies both because of (unpredictable) variations in schools catchment areas and because of the overall increase in the number of schools offering triple science. My first stage equation is:

$$D_{ipt} = \theta_1 z_{pt}^2 + \theta_3 X_{ipt} + \theta_t + \theta_p + v_{ipt} \tag{7}$$

where z_{pt}^2 is my instrument: the share of schools reachable in year t by student i, residing in block p, that were offering triple science in year t, when i applied to secondary school. X_{ipt} are neighbourhood and individual controls[24] and θ_t and θ_p are cohort and neighbourhood fixed effects respectively; v_{ipt} is the error term.

This instrument compares students attending schools that offer triple science with students attending schools not offering it, i.e. it uses across school within neighbourhood variation. However, offering triple science is likely to be related with other school characteristics, like school quality, that may directly affect the choices of majors at university. This issue may be more relevant when we use across school rather than within school variation because differences in quality across schools are likely to be more sizeable than differences within schools over time. Section 1.4.5 addresses this concern by including as control the average quality level of 'reachable' schools in each year catchment areas.

[21]With some exceptions for students with siblings attending the same school or for students with special education needs. Since I do not have the full set of information necessary to simulate the exact admission formula for each school, I can't adopt an RDD strategy.

[22]In order to exclude exceptions I eliminated outliers (the distances higher than the 0.01 percentile for every school.

[23]In total there are more than 30,000 LLOAs in England and Wales and each LLOA contains on average 1500 households.

[24]The average primary school grade in math, English, science of students in the neighbourhood, the share of girls and of FSME students in the neighbourhood, beside the usual individual controls.

1.4 Results

This section begins by showing results obtained with the first instrument. It first shows the overall effect of taking more science in high school in term of post-16 outcomes (Subsection 1.4.1). Second, it describes who decides to take triple science, when exposed to the option of taking it, by characterizing compliers (Subsection 1.4.2). Third it analyses how the effect is heterogeneous, depending on gender, ability and socio-economic status (Subsection 1.4.3). Finally it checks the identifying assumptions and whether the main findings are robust to the second identification strategy (Subsections 1.4.4 and 1.4.5).

1.4.1 Main Results

Table 2 presents the main estimates of the effect of taking triple science at age 14 on the probability of choosing at least one natural science subject at age 16 (KS5) and a STEM major at university.[25] The Table proceeds by estimating the effect of interest under different specifications. Column 1 displays results from a simple OLS regression, in column 2 I add school fixed effects, column 3 follows Altonji (1995) and uses as instrument for triple science the share of students taking triple science in school s and year t. Column 4 uses my first instrument (z_{st}), but instead of including school fixed effects only adds some school level controls.[26] Column 5 shows results from my preferred specification: it estimates equation 5, using the identification strategy whose first stage is described by equation 6, that includes school fixed effects. Finally column 6 includes a school-specific trend. Reassuringly, the coefficients of columns 5 and 6 are very similar, suggesting that schools offering triple science are on a similar trend. Column 7 estimates the specification of equation 5, but it eliminates controls (X_{ist}), to check whether the instrument z_{st} is correlated with observable cohort-specific characteristics. Again, the coefficients of columns 5 and 7 are very similar, suggesting that -conditional on my fixed effects- the instrument is quasi randomly assigned. As expected the bias in the OLS estimates is upward: the coefficient indeed gets smaller as we correct for all different layers of selection. The Table shows that, if a student strengthen her science preparation at age 14, she is 5 percentage points more likely to take science at age 16 and

[25] I adopt throughout the paper a narrow definition of STEM majors, that includes engineering, math and the three natural science subjects. The dependent variables in all cases are dummies equal to one if students attend each different course and equal to 0 if they do not attend those courses or do not continue studying at all.

[26] In particular, the share of girls attending school s in year t and the share of FSME (Free School Meal Eligible).

1.5 percentage point more likely to choose a STEM major at university.

Table 3 shows the coefficients obtained from estimating equation 6 on some other age 14 (KS4), age 16 (KS5) and university outcomes. The top panel shows results on KS4 grades and on the number of exams taken in KS4 and KS5. Since triple science is more difficult, taking it reduces the average science grade at KS4. Columns 2 and 3 show instead that there are not spillovers on other subjects' grades. Columns 4 and 5 investigate whether the total number of qualifications taken at age 14 and 16 changes, as a consequence of the new course offered. The results show that the number of exams taken at age 14 slightly increases.
The second panel refers to outcomes at age 18, the results of KS5 exams. Column 1 shows that the policy does not have any effect on the probability of continuing to study at age 16. This is probably because the instrument mainly affects high ability students, who would continue to study in any case. This is a very important result, because a change in the probability of enrolling in science subjects at age 16 may be driven by both a change in the likelihood of continuing to study after age 16 and a change in the likelihood of choosing science subjects, conditional on continuing. Column 1 tells that the coefficient estimated on KS5 subjects comes entirely from an increase in the second component, because the first is not affected by the policy. The result displayed in column 2 shows that the effect of studying triple science is not only limited to the pure natural science subjects but it also has spillovers on math, for instance. The third panel refers to outcomes at university. Column 1 shows again that the policy does not have any effect on the probability of continuing to study at university.[27] The other columns show the effect on choices of majors and on the quality of the institution attended. Students taking triple science are more likely to attend institution belonging to the Russell group.[28] Moreover studying more science in high school increases not only the probability of enrolling in STEM majors but also, even more importantly, the probability of graduating on time in these majors.[29] This is extremely relevant given the large debate that is taking place in many countries, and the US in particular, about the low persistence of students in scientific fields (Arcidiacono et al., forthcoming; Stinebrickner and Stinebrickner, 2014).

One important aspect, which due to lack of available data is largely under-explored in the literature, is the extent and the presence of subjects complementarity and substitutability. If one takes more science at age 14, is she more likely to take other (complement) subjects

[27] Note that even if the magnitude of the coefficient is similar to the other coefficients, the baseline in this case in much larger: the average is 36% in this case.

[28] The Russell group represents 24 leading UK universities in terms of research and teaching.

[29] The results on university outcomes are estimated on students taking the final KS4 exam in the years 2005-2007 only, otherwise there is no information on whether the students graduated from university.

in the meantime and, more importantly, from which (substitute) subjects does she opt out? Moreover, does more science in high school lead just to more science later on or does it also trigger a virtuous cycle where students start studying more challenging and difficult subjects, even if not explicitly related to the three natural sciences (such as math, engineering, medicine...)? Table 16 in the Appendix shows the coefficients and standard errors obtained from estimating equation 5 using each time a different KS4 subject as dependent variable.[30] Tables 17 and 18 report the same type of estimates but with respect to KS5 subjects and university majors, respectively. Columns 1 and 2 refer to the entire sample, columns 3 and 4 to girls and columns 5 and 6 to boys. Students who take triple science at KS4 tend to drop more vocational subjects, some foreign languages like German and some other core subjects like history. In terms of KS5 courses, taking triple science induces students to take more natural science subjects and more math later on, and to drop more vocational subjects, like media and accounting. Finally, triple science increases the probability of choosing scientific subjects at university, like physics, engineering and medicine, but also non scientific but difficult subjects, like classical languages. It decreases, instead, the probability of enrolling in law and architecture. The effect masks substantial gender heterogeneity: the findings on STEM majors are entirely driven by boys. Subsection 1.4.3 further explores this aspect.

It is, however, difficult to draw conclusions from the coefficients of Tables 16, 17 and 18. Anecdotic evidence may suggest that a vocational course in music is very different from an advanced course in science at age 14, but to evaluate each subject along some objective dimensions, Table 4 uses a more formal procedure. I define courses along two dimensions: (i) 'high achievers' courses, characterized by a high average primary school grade of students choosing them in out-of-sample academic years; (ii) 'female dominated' courses, characterized by a high share of girls attending the courses in out-of-sample academic years (2002-2005). Figure 5 describes each subject, along these dimensions. In particular it shows three scatter-plots where for each course is displayed on the x-axis the share of girls usually enrolled in it and on the y-axis the average primary school grade of student attending it. Triple science is, by far, the course at KS4 that is attended by the best students, followed by foreign languages, history and geography. For what concern KS5 options, math is the most challenging course, followed by physics, chemistry and foreign languages. For university majors, medicine, languages and STEM subjects are attended by very good students while education, subjects allied to medicine and art are attended by the worst students on average. The correlation between the ability of students usually attending each course and the share of girls enrolled in those courses is negative. This is surprising, given that on average girls have higher grades

[30] I exclude math and English because compulsory in KS4.

than boys in primary school.

Table 4 shows whether students start choosing more 'high achievers' courses at age 18 (KS5) and at university as a consequence of taking more science at KS4.[31] Taking advanced science at age 14 induces students to choose more difficult subjects later on. Students taking triple science are induced to choose courses at age 16 whose average grade in primary school of students usually attending is about 0.2 standard deviations higher. Moreover, for KS5, I disentangle how much of the reported increase is automatically due to the higher probability of choosing natural science subjects and how much to the fact that students choose other (complement) more 'high achievers' subjects, different from the three natural sciences. I find that the increase is partly driven by an higher probability of choosing science courses (63%) and partly due to a higher willingness to enroll in other difficult subjects not strictly in the natural science field (37%).[32] The same is true for university majors, but the magnitude of the effect is smaller.

1.4.2 Compliers' characterization

This Section analyses who decides to take triple science, when the school offers it. This is useful to understand, on one side, how students make decisions about which subject to take at age 14 and, on the other side, whether the heterogeneity in the coefficient β_1 is actually driven by differences in the actual treatment effect or, since I estimate local effects, by differences in compliers' type. Even if teachers in the UK usually make recommendations to students about which field courses to choose, the choice of whether to actually take triple science or not is a free decision made by students.[33]

Students will choose to take triple science if their expected utility when $D = 1$ is higher than their expected utility when $D = 0$. This may be because triple science reduces their costs (or their perception of the cost) of graduating in certain majors or of graduating at all or because triple science directly increases their productivity, and therefore wage, in more scientific jobs. The contribution in terms of utility of taking triple science with respect to the second best

[31] To obtain these results I multiply the coefficients displayed in Tables 16, 17 and 18 by the numbers displayed in Figure 5 and I sum the series. Standard errors are computed through the Delta method.

[32] These results are available upon request.

[33] One caveat should be considered when interpreting the results: sometimes supply of triple science is constrained since classes in the UK cannot be larger than 30. Since schools mainly prioritize based on previous science and math scores, any differences in the probability of taking triple science based on previous test scores may be taken with caution. It may not be driven by students' willingness to take triple science, but by schools admission rules.

option, will not be the same for all students: there are some students who may not find as beneficial to take triple science, like students already very good in science or students with very strong preferences towards other subjects.[34] This means that the likelihood of taking triple science will not be the same for everybody: it will depend on students preferences, on their innate ability and on their perceptions towards their abilities.

The first row of Table 5 shows the first stage regression for the entire sample. Being unexpectedly exposed to the triple science course increases students' probability of being enrolled in triple science by 15 percentage points. The F statistics is around 2800.
Table 5 then characterizes compliers for the entire population and separately for boys and girls (columns 2 and 3, respectively). I obtain information on compliers' characteristics looking at the first stage for several covariate groups. For instance the ratio between the instrument's coefficient of the first stage estimated on the sample of females only (0.149) and the coefficient of the first stage estimated on the entire sample (0.163) represents the relative likelihood that a complier is female.[35] Table 5 displays coefficients from estimating equation 6, the first stage, on different subgroups of the population. The first column refers to the entire sample and splits it in different covariates groups. It shows that compliers are more likely to be very good students in primary school: the relative likelihood a complier is in the top 20th percentile of test scores in primary school is more than two. Moreover compliers tend to be high income students and, interestingly, there does not seem to be any particular gender difference in compliance. The second and the third columns compare compliers for the subgroups of girls and boys respectively and show that compliers' characteristics are very similar between these two groups.

1.4.3 Heterogeneity

This section evaluates the heterogeneity of the effect of strengthening the science curriculum in high school for different subgroups of the population. In particular, I analyse the heterogeneity of the effect by gender, socio-economic status and previous grades in science.

The first panel of Table 6 looks at whether attending more science classes at high school has a different effect depending on students science grades in primary school. In particular the

[34] Unless triple science has a positive effect also in reducing the cost of taking exams in other subjects, for instance through changes in self confidence.

[35] First stages in this case do not include any control a part from year and school fixed effects. This does not affect the effect of interest because controls are not correlated with the instrument.

Table looks at the probability of enrolling in STEM majors and of persisting in these studies. The group mostly affected by the policy are the middle-high ability students. The very high ability students would probably be very well prepared in any case and are less likely to be at the margin, the low ability students are instead less likely to be affected by the policy at all.

The second panel analyses heterogeneity by socio-economic status (SES).[36] The effect on science at age 16 is slightly stronger for low SES students, the effect on university outcomes is instead more difficult to estimate with enough precision for low SES students because of the small share (20%) of low SES students attending university.

The third panel analyses gender heterogeneity. The effect is positive for both boys and girls, but the effect on STEM majors is entirely driven by boys. Still, girls are affected by the policy: they are induced to enrol in more scientific majors, but tend to choose more female-dominated science majors like medicine instead of engineering.

Table 7 summarizes the results on gender, following the same method adopted in Table 4 but it looks at the sample of boys and girls separately. While the first row shows that girls tend to choose more challenging subjects (i.e. more 'high achievers' subjects) in about the same proportion as boys, the second row shows that they still opt for female-dominated subjects (like medicine for instance).

1.4.4 Checks to the identification strategy

As stated in Section 1.3, the instrument used in the analysis relies on some assumptions.

First, the assumption that the information set of both the treatment and the control groups of students at age 11 does not include the information on whether the schools that do not offer triple science when students apply are going to offer triple science in three years. To check this I include all schools in the sample (both offering and not offering triple science when student i applies) and I estimate the following equation:

$$W_{ist} = \alpha_1 z_{st}^{11} + \alpha_2 z_{st} + \alpha_3 X_{ist} + \xi_s + \xi_t + \eta_{ist} \tag{8}$$

[36]Two separate proxies of socio economic status are available in the NPD: Free School Meal eligibility (FSM), a dichotomous variable indicating whether the student is eligible for or in receipt of FSM (approximately 14% of students) and Income Deprivation Affecting Children Index (IDACI), that indicates the proportion of children under age 16 in the local area where the student lives who are living in low income households (the median is 16% of low income households in the area). Table 6 uses only FSM, but results are consistent using the other proxy.

where W_{ist} are several outcomes (like the dummy for whether student i chooses a STEM major or whether he graduates in it) or pre-determined characteristics (like the average science grade in high school, his gender etc); z_{st}^{11} is a dummy equal to 1 if school s attended by student i in cohort t offered triple science when the student was 11 and chose her high school and z_{st} is my usual instrumental variable. Table 8 shows the results with (panel 1) and without (panel 2) school specific trends. The coefficient α_1 is not significant for most variables and in any case is usually extremely small, suggesting that, even if parents and students know, when they choose their school, whether $S = 1$ or 0, they do not select schools correspondingly- at least in terms of observable characteristics. This is consistent with the notion that students cannot freely choose their schools because schools, when oversubscribed, have to select students based on geographical distance.

Second, the assumption that schools decide when to start offering triple science not based on the quality of the current cohort attending the school. Table 9 provides evidence that, when using my identification strategy, the timing of the introduction of the triple science option is not correlated with (observable) characteristics of current students in the school. The Table runs a set of placebo tests, where I estimate equation 6 (without controlling for X_{ist}) and where the dependent variable is a pre-determined outcome, the grade in the science course in primary school, and should not be correlated with the instrument. Therefore the triple science dummy (TS) should not be significant, unless the instrument is not taking full care of selection. The Table has the same structure of Table 2 and it shows how different identification strategies may fail to address endogeneity. Column 1 shows results from a simple OLS regression, column 2 adds schools fixed effect, column 3 replicates the specification used by Altonji (1995) and uses as instrument for triple science the share of students taking triple science in school s and year t, Column 4 uses my instrument but instead of including schools fixed effects adds some school level controls.[37] column 5 refers to my preferred specification and it includes also school fixed effects. Reassuringly, the effect in this case is 0. Finally column 6 includes school specific time trends, and the coefficient is again 0. Table 15 in the Appendix shows results from a set of balancing tests obtained estimating the same specifications as in columns 5 and 6 for a bunch of other predetermined observable characteristics. All balancing tests show that the treatment is not correlated with observable characteristics of the current students in the school.

Moreover, I check for the presence of parallel trends. In particular, I check whether, before school s started to offer triple science, the trend was parallel to that of all other schools still

[37]This column partly replicates, even if in a very different context, Joensen and Nielsen (2015)

not offering triple science, i.e. the school did not start offering triple science because it was already on a different trend. To do this, I augment my reduced form regression with leads and lags of the instrument (following Autor (2003)):

$$y_{ist} = \sum_{t=0}^{m} \gamma_{\tau-t} z_{s(\tau-t)} + \sum_{t=0}^{q} \gamma_{\tau+t} z_{s(\tau+t)} + \zeta_t + \zeta_s + u_{ist} \tag{9}$$

where z_{st} is my instrument, τ is the year school s starts offering triple science, ζ_s and ζ_t are the usual school and year fixed effects and u_{ist} is the error term. I then check for the presence of parallel pre-treatment trends by evaluating whether all coefficients $\gamma_{\tau-t}$ are close to 0, for every τ. Figure 6 shows that the trends are parallel before the introduction of the advanced science course and there is a jump in the outcomes and in the treatment correspondingly exactly to the year of the introduction of the new course.[38] This confirms the results obtained in Table 8 and 9.

Another possible concern is that, once a school sets up all arrangements in terms of teaching qualifications and staff in order to offer triple science, it may start to offer more science courses at KS5 as well. In the UK about 60% of the schools offer both KS4 (age 14) and KS5 (age 16) exams. This would imply that part of the effect I find may be purely mechanical: students take more KS5 science courses because the set of options changes also at KS5. I address this concern in Table 10. Columns 1 and 2 look at how the probability of offering science at KS5 evolves over time and whether it corresponds exactly to the cohort when the school starts offering triple science at KS4. The correlation is 0: the probability of offering science at KS4 is not correlated with the probability of offering science at KS5. Columns 3 and 4 look at whether the effect of studying triple science on the probability of choosing science at KS5 is larger for schools offering both KS4 and KS5 courses than for schools offering KS4 courses only. The effect is identical. If part of the effect I find in my results was mechanical, it would be stronger for schools offering both KS4 and KS5 exams.

Moreover, one may worry that taking triple science could potentially directly affect the possibility of being admitted to STEM majors at university and to science courses at KS5, especially if there is a limited amount of available places. This would imply that my results are not generated by a higher number of applications but just by a higher probability of being admitted, given the same choices. Still, while universities often require some A level subjects in order to admit students to certain majors, in no case they require specific KS4

[38] I also estimated the same graphs but using predetermined characteristics as dependent variables: in this case there is no jump at year 0, nor at year -3, that correspond to the time when students know, when applying, that the school offers triple science. These results are available upon request.

subjects. For instance, in 2013, a KS5 exam in math was required in 13% of the cases (i.e. of major-university combinations) and at least one KS5 exam in science was required in 12% of the cases. In no case, in 2013, there was a specific requirement for age 14 (KS4) subjects.[39]

Finally, it may be that the simple fact of having the possibility of being enrolled in advanced science but having been excluded, for example because the class was oversubscribed and schools had to select students, may generate a direct effect on some students and may therefore violate the exclusion restriction assumption. This is impossible to test. Table 11 however exploits some of the institutional features of the UK scholl system to evaluate how problematic this may be. Figure 7 plots the distribution of the size of triple science courses in each school. From the Figure it is clear that class size bunches at multiples of 30. There is a discontinuity both corresponding to 30 students and corresponding to 60 students. Since class size in the UK is required to be lower than 30, this Figure suggests that in some cases the triple science course was oversubscribed, and schools had to select students. Unfortunately the exact admission rule is different for each school and is not publicly available. Table 11 exploits this feature of the system and runs the main specification (using as first stage equation 6) on the sample of schools where the triple science course was most likely not oversubscribed, because the number of enrolled students was not close to the maximum.[40] The results of this exercise are very similar to the main ones.

1.4.5 Second instrument

Table 12 shows the results obtained from the second identification strategy, described in Section 1.3.3.[41] The first column does not include neighbourhood fixed effects, but controls for the lagged value of my instrument: it compares neighbourhoods which had the same share of reachable schools offering the triple science course the previous year and it exploits variation between t and $t-1$. The second column includes neighbourhood fixed effects.

This instrument compares students attending different schools which offer or do not offer triple science, i.e. it exploits across school within neighbourhood variation. However, the

[39] Data are taken from http://www.thecompleteuniversityguide.co.uk/courses/search

[40] Those schools where the number of students enrolled in the triple science classes was not between 28 and 32 or between 58 and 62.

[41] Since there is no information on postcode in primary school for students who finished high school in the years before 2007, this section only refers to the years 2007-2010. For these cohorts, however, I have information on whether they graduated only for the students who took KS4 exams in the year 2007, so I only analyze effects on enrollment and on KS5 outcomes.

probability of offering triple science is likely to be related to other school characteristics, like school quality, that may directly affect the choices of majors at university. I address this point in Column 3, where I include the average quality of the set of reachable schools in year t as a control. I proxy school quality using the school value added in the previous years (2002-2005). The results confirm the robustness of the first identification strategy: the estimated effects are very similar to those found in Table 3.

1.5 Mechanisms

This Section digs into the mechanisms that may generate the effect found in Section 1.4 and explores whether the effect obtained is actually generated by changes in curriculum or, since the treatment has multiple components, it is actually driven by changes in the peer composition of the courses attended or in the type of teachers in the school.

1.5.1 Peers

First, I analyse the peers channel. In particular, I use the following measure of peer quality in science (Q_{ist}) for student i, attending school s in year t who takes science courses D_{ist}:[42]

$$Q_{ist} = \overline{X}^{D}_{(-i)st} \tag{10}$$

where $\overline{X}^{D}_{(-i)st}$ is the average science grade in primary school of students taking age 14 science course D[43], in school s in year t (excluding i).

The first panel of Figure 8 shows how peers' composition in the science course taken at age 14 changes for schools offering triple science or not. The dashed line plots the density of Q_{ist} in the age 14 science course for students attending schools not offering triple science. The solid line refers instead to schools offering triple science. The figure shows that when schools offer triple science there is a concentration of very high ability students able to attend the science class with peers of much higher quality than before. Column 1 of Table 13 confirms this finding: it shows how peers' quality in science courses changes after the school starts

[42] D_{ist} takes a different value if the student takes triple science, double or single science.

[43] Since there is no information about the exact class but only about the type of science course, I use the average grade in primary school of students taking the same course.

offering the advanced science course, depending on students' primary school grade in science. The quality of peers in the science class decreases for lower ability students and increases quite extensively for higher ability students.

To control for this dimension and check whether the effect found in Table 3 comes mostly from changes in the peer composition or from changes n the curriculum, I control for peer quality in equation 5. Since students self-select into different types of science course at age 14, peers' quality may be endogenous. I therefore instrument peer quality by using within-school over-time changes in peers' composition (following Hoxby (2000)). In particular, I use the fact that classes in the UK cannot be larger than 30 (as shown in Figure 7).[44] I therefore predict, based on predetermined characteristics,[45] the probability of getting into triple science and I take the average science grade in primary school of the 30 or 60 students (depending on the number of triple science classes offered) with the highest probability of being enrolled into triple science. I then exploit within school over time variation in the average quality of these students and of all other students in school s and year t, allowing the effect to be different depending on whether the school offers (unexpectedly) triple science. My first stage equation is:

$$Q_{ist} = \theta_1 z_{st} + \theta_2 \widehat{Q^{top30}_{st(-i)}} + \theta_3 \widehat{Q^{others}_{st(-i)}} + \theta_4 \widehat{Q^{top30}_{st(-i)}} * z_{st} + \theta_5 \widehat{Q^{others}_{st(-i)}} * z_{st} + \theta_5 X_{ist} + \theta_s + \theta_t + \eta_{ist} \quad (11)$$

where z_{st} is the first instrument - the dummy equal to 1 student i was unexpectedly exposed to the option of choosing triple science- $\widehat{Q^{top30}_{st(-i)}}$ is the average science grade in primary school of the 30 (or 60) students with the highest predicted probability of being enrolled in triple science and $\widehat{Q^{others}_{st(-i)}}$ is the average science grade in primary school of all other students; θ_s and θ_t are school and year fixed effects and η_{istj} is the error term. Panel b of Figure 8 shows how the instrument works. The solid line refers to the average science grade in primary school for students predicted to attend the triple science class, the dashed line refers to all other students.

Table 13 displays the results. Columns 2 to 6 show that the effect of triple science is very similar to what found before, even after controlling for changes in peers' quality. The joint F statistic is 35.

[44] While for primary schools this requirement is compulsory, it is just recommended for secondary school.

[45] KS2 and KS3 science grades (both teacher assessed and from standardized exams) , gender, Free School Meal Eligibility.

1.5.2 Teachers

Unfortunately, there are no data on teachers in the UK. The only information available refers to the yearly number of teachers and of qualified teachers in each school. Table 14 shows that neither the overall number of teachers nor the number of qualified teachers changed significantly once the school introduced the triple science option. This suggests that teachers' quality and quantity did not increase in correspondence to the introduction of the advanced science course.

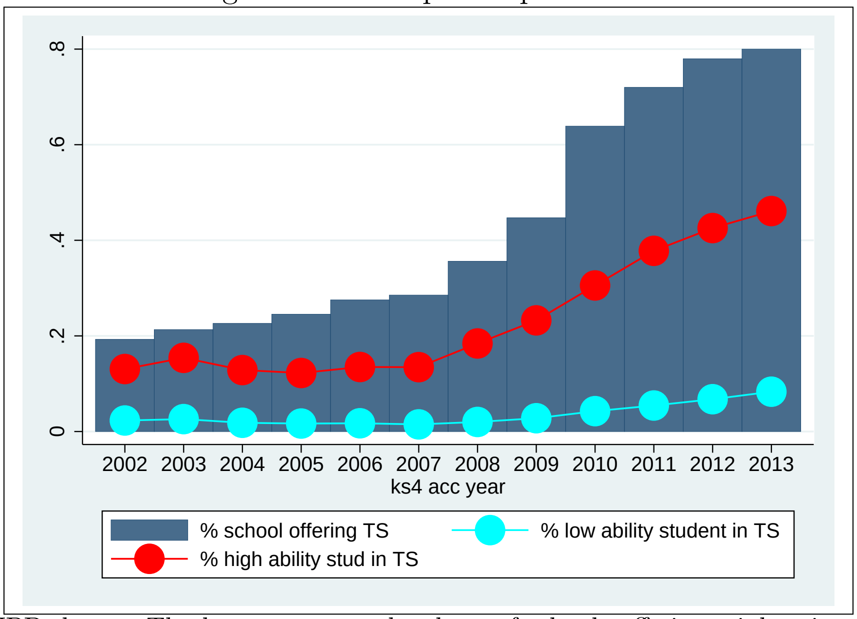
1.6 Conclusions

This paper uses a reform that increased the supply of advanced science courses in high school in the UK to ask whether high school curriculum affects post-16 outcomes, in particular the probability of graduating and of graduating in a STEM major. Moreover it asks whether the effect is heterogeneous with respect to gender and socio-economic status.
My estimates suggest that offering more science in high school improves educational outcomes in many domains. Advanced science course in high school has no clear effect on high school graduation and university enrolment on average, probably because they mostly affect students who would continue studying in any case. Moreover it shifts very high ability students towards high quality universities. More science at age 14 significantly increases the probability of enrolling and, very importantly, of graduating into a STEM major and scientific majors at university. This effect masks however a substantial and interesting gender heterogeneity. When exposed to the option of studying more science in high school at age 14, there is no gender difference in the likelihood of taking more science. However, the difference arises later on, at university, when subject choices are likely to be correlated to occupation and jobs: both boys and girls opt to more challenging and more scientific courses on average, but girls choose more female-dominated subjects like medicine, instead of engineering and math.
Since the advanced science courses attracts a very favourable selection of peers, I disentangle how much of the effect is driven by peer composition and I show that different peers' quality is not the main driver behind my results: what matters is the actual change in the curriculum at high school.

My findings show that there is a certain degree of persistence between what is studied at high school and what is studied at university. A optimal design of the high school curricula may be useful to improve the matching between supply and demand of specific skills.

Figures

Figure 1: Take up in triple science



Source: NPD dataset.The bars represent the share of schools offering triple science; the red dots represent the share of high ability (based on English, math and science primary school grade, top 40 %) students taking triple science and the blue dots show the share of low ability (based on primary school grades, bottom 60 %) students taking triple science, by year.

Figure 2: Timeline of the English educational system

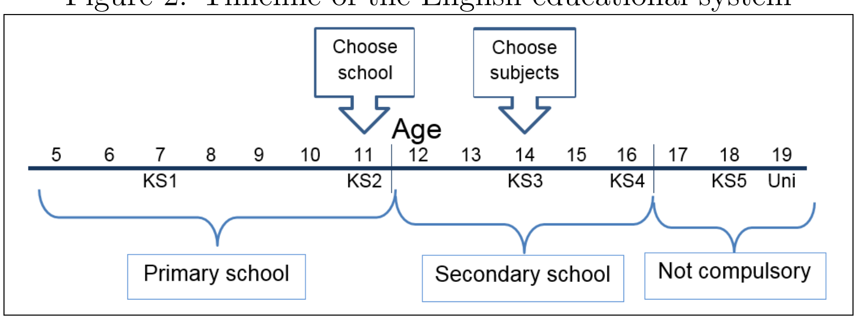
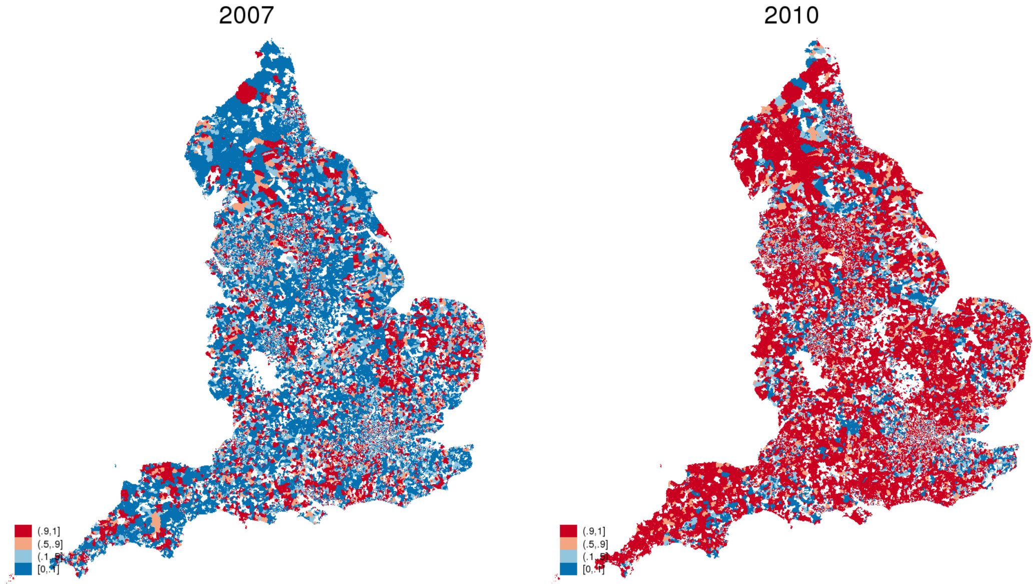


Figure 3: Spacial variation in the diffusion of triple science, over time

offering in 2005 (exam taken in 2007) | offering in 2008



Source: NPD dataset. The two maps show the diffusion over time of the share of schools offering triple science, by LLOAs (Lower Level Output Areas).

Figure 4: Second instrument

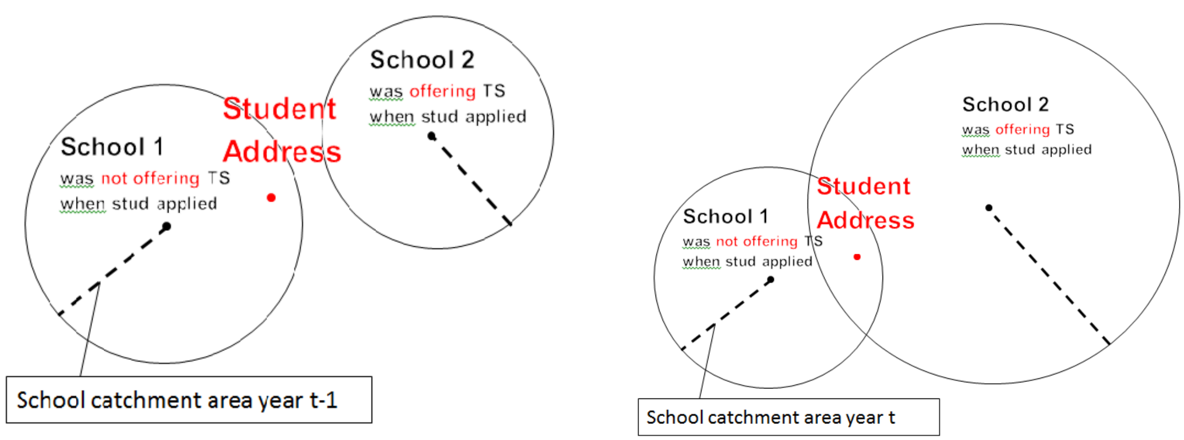
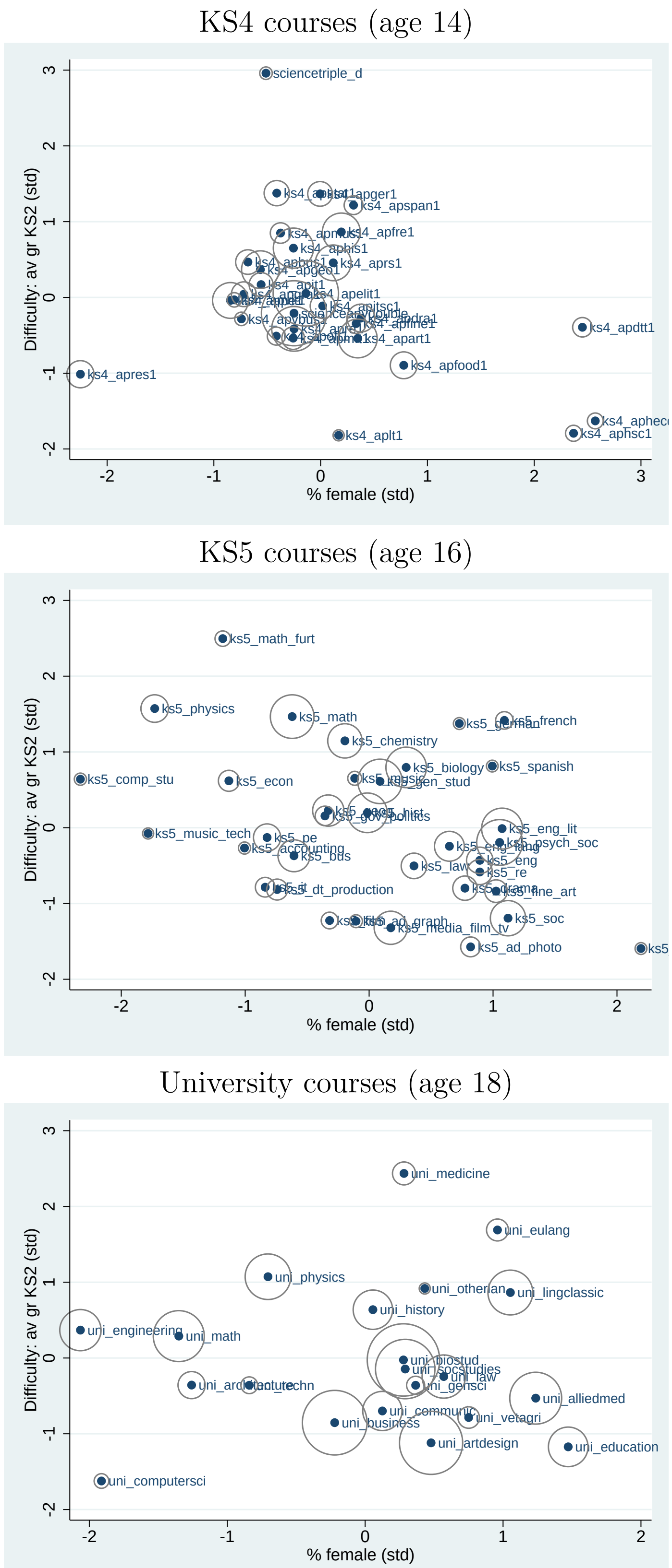
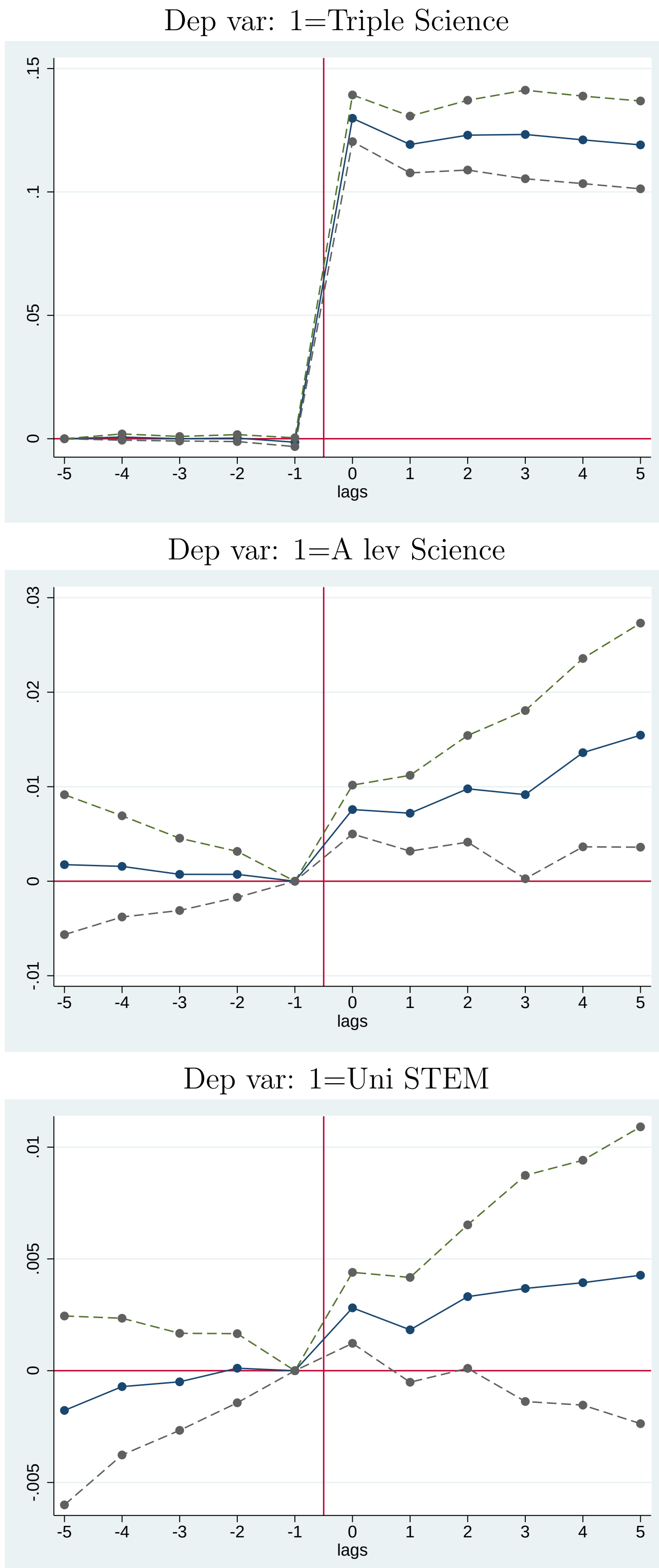


Figure 5: Subject descriptives



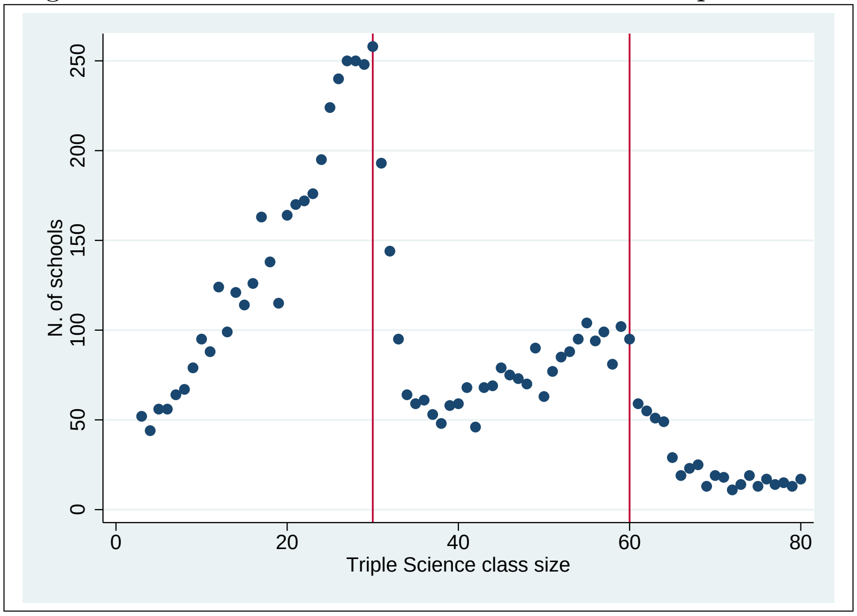
Source: NPD dataset. Subjects are described along two dimensions: the average primary school grade (in English, math and science) of students taking the course in out of sample years and the share of girls taking the course in out of sample years. The circles around each observation represent the number of students attending these courses.

Figure 6: Parallel Trends: Leads and Lags of the instrument



Source: NPD dataset. The continuous line represent coefficients, the dashed lines the 5% confidence intervals, obtained from estimating equation 9. Omitted category: one year before the treatment.

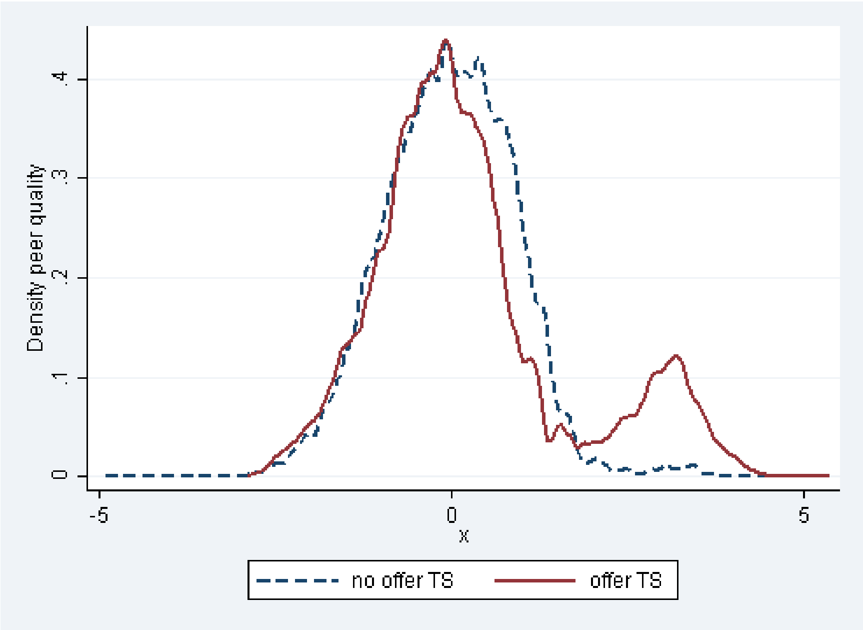
Figure 7: Class size and number of students in triple science



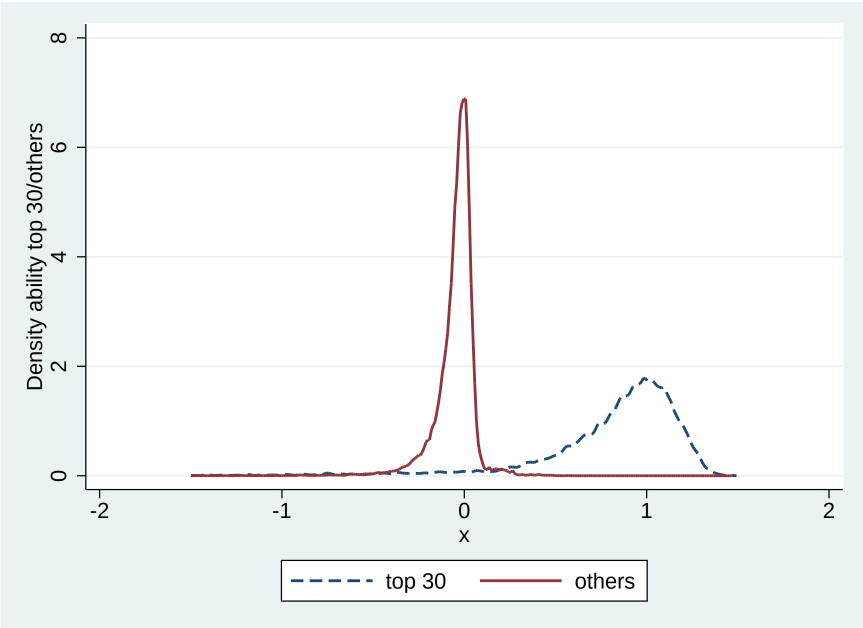
Source: NPD dataset. The dots are the number of schools, by triple science class size .

Figure 8: Peers

Actual peers' quality



Instrument



Source: NPD dataset. The first panel plots the distribution of science peers' quality, distinguishing whether the school offers triple science or not. The second panel plots the average peers quality for students predicted to take the TS class and students not predicted to take the TS class.

Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.
	Key Stage 4	
offer TS (unexpected)	0.196	0.397
1=Triple Sci	0.076	0.264
1=Double Sci	0.764	0.425
1=Single Sci	0.163	0.369
	Key Stage 5	
1=KS5 science (if KS5)	0.198	0.282
1=KS5 math (if KS5)	0.142	0.252
	University	
1=uni	0.348	0.470
1=STEM[a]	0.126	0.198
1=Russell	0.046	0.211
1=graduate[a]	0.481	0.361
	Demographics	
1=female	0.497	0.500
1=FSM eligible[b]	0.144	0.356

The summary statistics reported in the Table refer to the entire sample of students taking their final KS4 exams (at age 16) between 2005 and 2010.

[a] Conditional on going to university.

[b] Free School Meal Eligible.

Table 2: Results for science at age 17 and 19

	OLS	OLS-Fe	Altonji	IV	IV-Fe	IV-Fe tr	IV-Fe
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Dep var:				1=KS5 Science			
1=TS	0.334***	0.257***	0.147***	0.072***	0.051***	0.048***	0.054***
	(0.005)	(0.005)	(0.014)	(0.010)	(0.006)	(0.008)	(0.006)
1=female		-0.009***	-0.004***	-0.011***	-0.010***	-0.010***	
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
I sch gr sci		0.020***	0.019***	0.019***	0.021***	0.022***	
		(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	
N	1690451	1690451	1690451	1690451	1690451	1690451	1690451
Fstat			559372	2234	2065	1742	2066
Dep var:				1=STEM university			
1=TS	0.104***	0.072***	0.039***	0.024***	0.014***	0.012**	0.015***
	(0.002)	(0.002)	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)
1=female		-0.034***	-0.034***	-0.035***	-0.034***	-0.034***	
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
I sch gr sci		0.005***	0.005***	0.005***	0.006***	0.006***	
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
N	1690451	1690451	1690451	1690451	1690451	1690451	1690451
Fstat			559372	2234	2065	1742	2066
School Fe	No	Yes	No	No	Yes	Yes	Yes
School trends	No	No	No	No	No	Yes	No
School contr	No	No	Yes	Yes	No	No	No
Stud contr	No	Yes	Yes	Yes	Yes	Yes	No

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and english; schools controls: school size. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 3: Results for other outcomes

	[1]	[2]	[3]	[4]	[5]
Panel 1: KS4 (age 14) outcomes					
		Grades		N. Exams	
Dep var:	KS4 Eng gr[a]	KS4 Math gr[a]	Ks4 science gr	n exams ks4	n exams ks5[c]
1=TS	0.001	-0.026	-0.065**	0.438**	-0.021
	(0.031)	(0.028)	(0.027)	(0.210)	(0.022)
N	1332413	1339792	1690325	1690451	860615
ymean	0.022	0.021	0.000	10.303	3.416
Panel 2: KS5 (age 16) outcomes					
Dep var:	1=KS 5	1=KS5 math	1=KS5 Bio	1=KS5 Che	1=KS5 Phy
1=TS	-0.009	0.035***	0.037***	0.025***	0.024***
	(0.010)	(0.005)	(0.004)	(0.003)	(0.005)
N	1690451	1690451	1690451	1690451	1690451
ymean	0.509	0.056	0.040	0.026	0.065
Panel 3: University outcomes[b]					
Dep var:	1=uni	1=grad	1=Russell	1=uni med	1=grad STEM
1=TS	0.044*	0.041	0.022*	0.013**	0.033***
	(0.025)	(0.025)	(0.011)	(0.007)	(0.011)
N	966777	966777	966777	966777	966777
ymean	0.318	0.207	0.046	0.019	0.034

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and english; schools controls: school size. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Grades go from 0 to 7, but are standardized to have mean 0 and standard deviation 1.

[b] The results on university outcomes use only the 2005-2008 sample because otherwise there will be no information on the graduation outcomes.

Table 4: Summarizing effects on other subjects

	Δ ks5 courses	Δ uni major
High achievers	0.187***	0.021***
	(0.019)	(0.007)
Female-dominated	-0.050***	-0.008
	(0.017)	(0.008)

The coefficients are computed as $\sum_j \beta_j q_j$ where j indicates subjects, β_j is the subject specific coefficient estimated in Tables 17 and 18 and q_j is either 'high achievers'(the average primary school grade of taking the course j in out of sample academic years (2002-2005), standardized to have mean 0 and standard deviation 1) or 'female dominated' (the share of girls attending course j in out of sample academic years). Standard errors are computed through the delta method.

Table 5: Characterizing compliers

Sample	Everybody [1]	Only Girls [2]	Only Boys [3]
Panel 1: Entire Sample			
Z_{st}	0.175***	0.161***	0.188***
	(0.004)	(0.005)	(0.005)
N	1690451	849184	841267
Panel 2: Quintiles science grade in primary school			
	subgroup:	1st quintile av.	primary school grade
Z_{st}	0.009***	0.008***	0.009***
	(0.001)	(0.001)	(0.001)
N	339951	174093	165858
Ratio wrt tot FS	0.051	0.050	0.048
	subgroup:	2nd quintile av.	primary school grade
Z_{st}	0.038***	0.035***	0.041***
	(0.001)	(0.002)	(0.002)
N	341063	171845	169218
Ratio wrt tot FS	0.217	0.217	0.218
	subgroup:	3rd quintile av.	primary school grade
Z_{st}	0.099***	0.092***	0.105***
	(0.003)	(0.003)	(0.004)
N	336767	168450	168317
Ratio wrt tot FS	0.566	0.571	0.559
	subgroup:	4th quintile av.	primary school grade
Z_{st}	0.222***	0.208***	0.234***
	(0.005)	(0.006)	(0.006)
N	344551	171725	172826
Ratio wrt tot FS	1.269	1.292	1.245
	subgroup:	5th quintile av.	primary school grade
Z_{st}	0.449***	0.417***	0.479***
	(0.009)	(0.011)	(0.010)
N	328119	163071	165048
Ratio wrt tot FS	2.566	2.590	2.548
Panel 3: Socio-Economic Status			
	subgroup:	Low SES students	(yes FSM[a])
Z_{st}	0.084***	0.077***	0.092***
	(0.002)	(0.003)	(0.003)
N	223375	114446	108929
Ratio wrt tot FS	0.480	0.478	0.489

The Table reports results from the first stage for different subgroups of the population. Dependent variable: a dummy equal to 1 if the student takes triple science. Additional controls: year and school fixed effects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Free School Meal Eligible.

Table 6: Heterogeneity

Dep var:	1=KS5 sci	1=Russell	1=STEM	1=medicine	1=grad	1=grad STEM
	[1]	[2]	[3]	[4]	[5]	[6]
Panel 1: Quintiles science grade in primary school						
			3rd quintile			
1=TS	0.019	-0.002	-0.002	0.015	0.036	0.032
	(0.015)	(0.035)	(0.037)	(0.028)	(0.089)	(0.036)
N	336723	203148	203148	203148	203148	203148
ymean	0.045	0.024	0.026	0.017	0.188	0.023
			4th quintile			
1=TS	0.032***	0.041*	0.076***	0.017	0.084*	0.086***
	(0.010)	(0.021)	(0.021)	(0.014)	(0.046)	(0.019)
N	344500	197276	197276	197276	197276	197276
ymean	0.104	0.053	0.045	0.024	0.277	0.042
			5th quintile			
1=TS	0.053***	0.018	0.010	0.005	0.016	0.012
	(0.007)	(0.016)	(0.015)	(0.008)	(0.023)	(0.015)
N	328076	181689	181689	181689	181689	181689
ymean	0.254	0.146	0.097	0.040	0.414	0.090
Panel 2: Socio-Economics Status						
			High SES students (no FSM)			
1=TS	0.048***	0.024**	0.020	0.015*	0.037	0.033***
	(0.006)	(0.011)	(0.013)	(0.008)	(0.026)	(0.012)
N	1431595	818880	818880	818880	818880	818880
ymean	0.093	0.052	0.041	0.020	0.226	0.037
			Low SES students (yes FSM)			
1=TS	0.063***	-0.008	0.042	-0.003	0.100	0.024
	(0.018)	(0.044)	(0.039)	(0.035)	(0.090)	(0.036)
N	258804	147854	147854	147854	147854	147854
ymean	0.034	0.015	0.018	0.010	0.103	0.016
Panel 3: Gender						
			Girls			
1=TS	0.047***	0.027	0.003	0.023**	0.049	0.015
	(0.008)	(0.021)	(0.015)	(0.009)	(0.040)	(0.013)
N	849149	486068	486068	486068	486068	486068
ymean	0.080	0.053	0.020	0.030	0.239	0.019
			Boys			
1=TS	0.053***	0.018	0.037**	0.005	0.033	0.045***
	(0.007)	(0.013)	(0.017)	(0.006)	(0.029)	(0.016)
N	841234	480646	480646	480646	480646	480646
ymean	0.088	0.040	0.054	0.008	0.174	0.049

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and English; schools controls: school size. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 7: Summarizing effects on other subjects by gender

Dep var:	Δ ks5 courses		Δ uni major	
	Girls	Boys	Girls	Boys
High achievers	0.157***	0.211***	0.021**	0.028***
	(0.028)	(0.022)	(0.012)	(0.008)
Female-dominated	-0.025	-0.065***	0.014	-0.023**
	(0.026)	(0.020)	(0.011)	(0.010)

The coefficients are computed as $\sum_j \beta_j q_j$ where j indicates subjects, β_j is the subject specific coefficient estimated in Tables 17 and 18 and q_j is either 'high achievers'(the average primary school grade of taking the course j in out of sample academic years (2002-2005), standardized to have mean 0 and standard deviation 1) or 'female dominated' (the share of girls attending course j in out of sample academic years). Standard errors are computed through the delta method.

Table 8: Selection

	av KS2 gr[a]	sci KS2 gr[b]	1=FSM	1=KS5 sci	1=uni	1=STEM	1=grad STEM
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Z^{11}_{st}	-0.005	-0.008	0.002	0.005***	-0.002	0.001	0.001
	(0.005)	(0.006)	(0.002)	(0.001)	(0.004)	(0.002)	(0.002)
N	2882341	2882341	2882341	2882341	1468169	1468169	1468169
School fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School trend	No	No	No	No	No	No	No
Z^{11}_{st}	0.002	0.002	0.007**	0.004**	-0.003	0.001	0.001
	(0.006)	(0.002)	(0.003)	(0.002)	(0.005)	(0.002)	(0.002)
N	2285735	2285735	2285735	2285735	1309004	1309004	1309004
School fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Additional controls years dummies, school fixed effects. Robust standard errors clustered by school in parentheses. The dependent variables in column 4, 5 and 7 are set equal to 0 if students do not continue studying or if they do not take that subject. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] average grade in English, math and science.

[b] grade in science.

Table 9: Balancing Test

	OLS [1]	OLS-Fe [2]	Altonji [3]	IV [4]	IV-Fe [5]	IV-Fe tr [6]
Dep var:	1=Average Grade prim school[a]					
1=TS	0.927***	0.788***	0.802***	0.363***	0.042	0.045
	(0.013)	(0.015)	(0.054)	(0.052)	(0.026)	(0.034)
mfemale				0.232***		
				(0.053)		
mfsm				-1.545***		
				(0.051)		
N	1337202	1337202	1337202	1337202	1337202	1337202
School Fe	No	Yes	No	No	Yes	Yes
School time trends	No	No	No	No	No	Yes

Additional controls: years dummies. Robust standard errors clustered by school in parentheses.
* denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.
[a] Average grade in the KS4 exams in English, math and science.

Table 10: Robustness: offer KS5 Science

	Sch level regr (offer)		Stud in schools wo sixth form	
Dep var:	1=Offer KS5 Science	1=offer KS5 Math	All schools	only offer KS4
			Dep var: 1=KS5 Science	
	[1]	[2]	[3]	[4]
Z_{st}	0.002	-0.000		
	(0.004)	(0.004)		
1=TS			0.050***	0.053***
			(0.006)	(0.009)
N	5294	5294	1690451	751721
ymean	0.477	0.467	0.084	0.060

Column 1 and 2 are run at the school-year level. Columns 3 and 4 are run at the student level. Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and English; schools controls: school size. The dependent variables in columns 3, and 4 are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 11: Robustness: exclusion restriction

Dep var:	1=KS5 sci	1=Russell	1=STEM	1=medicine	1=grad	1=grad STEM
	[1]	[2]	[3]	[4]	[5]	[6]
1=TS	0.057***	0.024*	0.022	0.010	0.039	0.026**
	(0.007)	(0.014)	(0.013)	(0.009)	(0.028)	(0.012)
N	1613226	948058	948058	948058	948058	948058

The sample includes only schools where the triple science class is not likely to be oversubscribed (class size not around a multiple of 30). Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and English; schools controls: school size. The dependent variables in columns 3, 4, 5 and 6 are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 12: Identification based on the second instrument

	IV [1]	IV Neighb FE [2]	IV Neighb FE [3]
Dep. variable:		1=KS5 Physics	
1=TS	0.062***	0.065**	0.060**
	(0.008)	(0.026)	(0.028)
% reach school off TS_{t-1}	0.005***		
	(0.002)		
av. qual reach school			0.007
			(0.005)
Dep. variable:		1=KS5 Chemistry	
1=TS	0.041***	0.049	0.040
	(0.010)	(0.031)	(0.033)
% reach school off TS_{t-1}	0.001		
	(0.002)		
av. qual reach school			0.015**
			(0.006)
Dep. variable:		1=KS5 Biology	
1=TS	0.047***	0.045	0.032
	(0.012)	(0.037)	(0.038)
% reach school off TS_{t-1}	0.000		
	(0.002)		
av. qual reach school			0.019***
			(0.007)
Dep. variable:		1=Uni Engineering	
1=TS	0.006	0.034**	0.032**
	(0.005)	(0.015)	(0.016)
% reach school off TS_{t-1}	0.001		
	(0.001)		
av. qual reach school			0.003
			(0.003)
Dep. variable:		1=Uni Medicine	
1=TS	-0.010*	0.045**	0.044**
	(0.006)	(0.020)	(0.021)
% reach school off TS_{t-1}	-0.000		
	(0.001)		
av. qual reach school			0.002
			(0.004)
N	2860812	2861393	2861393
Neigh fe	No	Yes	Yes

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and english; neighbourhood controls: av grade in primary school in science, math, english, share of girls, share of low SES. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Robust standard errors clustered by neighbourhoodM in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 13: Peers

Dep var:	Q_{ist}[a]	1=KS5 sci	1=Russell	1=STEM	1=medic	1=grad	1=grad STEM
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Z*gr sci q1	-0.095***						
	(0.011)						
Z*gr sci q2	-0.060***						
	(0.008)						
Z*gr sci q3	-0.031***						
	(0.007)						
Z*gr sci q4	0.024***						
	(0.007)						
Z*gr sci q5	0.055***						
	(0.007)						
Z*gr sci q6	0.099***						
	(0.008)						
1=TS		0.053***	0.022**	0.024**	0.013*	0.042*	0.034***
		(0.006)	(0.011)	(0.012)	(0.008)	(0.025)	(0.011)
Q_{ist}		0.021***	0.018***	0.003	-0.001	0.014	0.004
		(0.005)	(0.004)	(0.004)	(0.003)	(0.009)	(0.004)
N	1648926	1621765	935630	935630	935630	935630	935630

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and english; schools controls: school size. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. Gr sci refers to sixtiles of the grade distribution in the science exam at the end of primary school (KS2). F statistic: 35.

[a] quality (based on science grade in ks2 (age 11) of peers in the same science class. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 14: Teachers

Dep. variable:	N teachers	N qualified teachers
	[1]	[2]
1=TS	1.604	1.577
	(1.267)	(1.249)
N	1022489	1022489
ymean	70.567	66.654

Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and english; schools controls: school size. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

1.7 Appendix

Table 15: Other balancing tests

	RF [1]	RF [2]	IV [3]	IV [4]
Dep var:	1=Grade English prim school			
Z_{st}	-0.000	0.005		
	(0.004)	(0.005)		
1=TS			-0.001	-0.002
			(0.023)	(0.023)
N	1690451	1690451	1690451	1690451
ymean	0.015	0.015	0.015	0.015
Dep var:	1=female			
Z_{st}	-0.002	-0.001		
	(0.001)	(0.002)		
1=TS			-0.009	-0.009
			(0.009)	(0.009)
N	1690451	1690451	1690451	1690451
ymean	0.502	0.502	0.502	0.502
Dep var:	1=FSM			
Z_{st}	-0.000	-0.000		
	(0.001)	(0.002)		
1=TS			-0.001	-0.001
			(0.008)	(0.008)
N	1690451	1690451	1690451	1690451
ymean	0.153	0.153	0.153	0.153
School Fe	Yes	Yes	Yes	Yes
School trend	No	Yes	No	Yes

Additional controls years dummies. All dependent variables are set equal to 0 if students do not continue studying or if they do not take that subject. Robust standard errors clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 16: Effect on other KS4 subjects (age 14)

Dep. var	All		Girls		Boys	
	[1]	[2]	[3]	[4]	[5]	[6]
	Coeff.	Se	Coeff.	Se	Coeff.	Se
English lit	0.068**	(0.030)	0.075**	(0.030)	0.061*	(0.032)
Statistics	0.011	(0.034)	0.010	(0.038)	0.011	(0.034)
DT food	-0.027*	(0.016)	-0.047**	(0.024)	-0.009	(0.013)
DT graphics	-0.015	(0.014)	-0.002	(0.017)	-0.027	(0.017)
DT material	-0.014	(0.014)	0.000	(0.011)	-0.024	(0.022)
Art design	-0.008	(0.019)	0.001	(0.025)	-0.015	(0.019)
History	-0.032*	(0.019)	-0.045*	(0.023)	-0.022	(0.021)
Geogr	0.007	(0.020)	0.010	(0.024)	0.005	(0.022)
French	-0.015	(0.028)	-0.010	(0.033)	-0.020	(0.027)
German	-0.065***	(0.018)	-0.072***	(0.022)	-0.060***	(0.018)
Business	-0.012	(0.019)	-0.012	(0.020)	-0.014	(0.021)
Drama	0.007	(0.014)	-0.001	(0.020)	0.013	(0.014)
Inf tech	-0.034	(0.031)	-0.020	(0.032)	-0.048	(0.035)
Music	-0.001	(0.008)	-0.012	(0.011)	0.009	(0.010)
Media	-0.012	(0.022)	-0.016	(0.025)	-0.009	(0.023)
Fine art	0.005	(0.014)	0.007	(0.019)	0.004	(0.013)
Office technology	0.016	(0.028)	0.008	(0.032)	0.022	(0.028)
Applied buss	-0.001	(0.014)	-0.004	(0.015)	0.000	(0.015)
Health care	0.003	(0.011)	0.009	(0.022)	-0.002	(0.004)
Applied IT	-0.009	(0.021)	-0.009	(0.021)	-0.008	(0.024)

Each line represents a different regression. Columns 1, 3 and 5 display the coefficients on the independent variable $1 = TS$. All dependent variables are set equal to 0 if students do not take that subject. Usual controls. Robust standard errors clustered at the school level. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 17: Effect on other KS5 subjects (age 16)

Dep. var	All		Girls		Boys	
	Coeff.	Se	Coeff.	Se	Coeff.	Se
Biology	0.035***	(0.005)	0.037***	(0.008)	0.034***	(0.006)
Chemistry	0.037***	(0.004)	0.032***	(0.006)	0.040***	(0.005)
Physics	0.025***	(0.003)	0.012***	(0.003)	0.036***	(0.005)
Math	0.024***	(0.005)	0.016**	(0.007)	0.031***	(0.007)
AD textile	-0.003*	(0.002)	-0.005	(0.003)	-0.001*	(0.000)
History	0.005	(0.005)	0.004	(0.008)	0.005	(0.006)
Economics	0.003	(0.003)	0.002	(0.003)	0.004	(0.005)
Law	-0.007**	(0.003)	-0.007	(0.005)	-0.008**	(0.004)
Psychology	-0.010*	(0.006)	-0.015	(0.011)	-0.006	(0.005)
Media film tv	-0.012***	(0.005)	-0.013*	(0.007)	-0.011**	(0.005)
German	-0.003**	(0.001)	-0.002	(0.002)	-0.003**	(0.001)
Music tech	-0.004***	(0.001)	-0.001	(0.001)	-0.008***	(0.002)
Accounting	-0.002*	(0.001)	-0.002	(0.002)	-0.002	(0.002)

Each line represents a different regression. Columns 1, 3 and 5 display the coefficients on the independent variable $1 = TS$. All dependent variables are set equal to 0 if students do not continue studying or if they do not take that subject. Usual controls. Robust standard errors clustered at the school level. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 18: Effect on other university majors (age 18)

Dep. variables	All		Girls		Boys	
	Coeff.	Se	Coeff.	Se	Coeff.	Se
Physics	0.006***	(0.002)	0.001	(0.003)	0.009***	(0.003)
Math	0.001	(0.002)	-0.002	(0.002)	0.003	(0.004)
Engineering	0.007***	(0.002)	0.003**	(0.001)	0.011***	(0.003)
Biology	-0.001	(0.003)	-0.001	(0.005)	-0.002	(0.004)
Veterinary agric	-0.001	(0.001)	-0.001	(0.002)	0.000	(0.001)
Computer sci	-0.001	(0.001)	-0.001	(0.001)	-0.000	(0.002)
Technology	-0.000	(0.001)	-0.000	(0.001)	-0.000	(0.001)
General science	-0.000	(0.001)	-0.001	(0.002)	0.000	(0.001)
Medicine	0.003*	(0.001)	0.006**	(0.002)	0.001	(0.001)
Allied medicine	0.004*	(0.002)	0.008*	(0.004)	0.000	(0.002)
Architecture	-0.003***	(0.001)	-0.002*	(0.001)	-0.004**	(0.002)
Other languages	0.000	(0.000)	-0.000	(0.001)	0.000	(0.001)
History	0.001	(0.002)	0.003	(0.003)	-0.001	(0.002)
Art design	-0.000	(0.003)	0.001	(0.005)	-0.002	(0.003)
Education	-0.001	(0.002)	-0.001	(0.004)	-0.001	(0.001)
Soc studies	0.003	(0.003)	0.005	(0.005)	0.001	(0.003)
Law	-0.004*	(0.002)	-0.006*	(0.003)	-0.002	(0.002)
Business	0.001	(0.003)	0.001	(0.004)	-0.000	(0.004)
Communication	0.000	(0.002)	0.001	(0.003)	-0.001	(0.002)
Ling classic	0.005**	(0.002)	0.004	(0.004)	0.006***	(0.002)
Eu languages	-0.000	(0.001)	-0.000	(0.002)	-0.000	(0.001)

Each line represents a different regression. Columns 1, 3 and 5 display the coefficients on the independent variable $1 = TS$. All dependent variables are set equal to 0 if students do not continue studying or if they do not take that subject. Usual controls. Robust standard errors clustered at the school level. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Chapter 2

Multi-task Agents and Incentives: the Case of Teaching and Research for University Professors

2.1 Introduction

The study of principal-agents relationships and the design of the optimal incentive provision systems have a long tradition in economics. A particularly complex and very common situation arises when agents have to allocate their time and effort among different tasks. In this case, the provision of incentives on one task only may distort multi-task agents' behaviour: individuals may respond by increasing effort in the activities subject to incentives, crowding out time and energy from other uses. This is especially the case if performance in other tasks is not easy to measure, and if there are no other reasons, such as social pressure or intrinsic motivation, to perform them in any case (Holmstrom and Milgrom, 1991, 1994; Brüggen and Moers, 2007; Fehr and Fischbacher; Bandiera et al., 2010; Benabou and Tirole, 2003; Prendergast, 2008).

While the theory related to multi-task agents is very well-developed, starting from the seminal work by Holmstrom and Milgrom (1991), empirical tests to the size and the sign of the behavioural responses predicted by this type of models are difficult to implement because of very heavy data requirements, first of all the need of an individual measure of performance for each task, which is often not easily observable. The empirical literature is therefore scarce and the actual economic cost of standard incentives for multitasks agents is still largely unknown. In practice, it depends on how the different tasks interact in the agent's production and cost functions and it is therefore specific to the actual tasks taken into consideration.

This paper analyses one of the leading examples of multi-task agents: the case of university professors. Faculty members allocate time among many activities, mostly teaching and research. Incentives in most countries are however strongly skewed towards research: the 'Publish or Perish' paradigm is the most popular criterion for faculty hiring and promotion decisions in universities. This paper analyses the overall consequences of strong research incentives on teaching and research outcomes. It evaluates, first, the direct impact of research incentives on research performance itself; second, it studies the indirect effect of research incentives on teaching quality. Moreover, in order to understand the overall impact on teaching and research performance at the university level, it analyses how the composition of professors in terms of teaching and research skills changes under an incentive scheme more skewed towards research. Finally, by analysing the correlation between teaching and research skills for each professor, this paper discusses what may be the costs and benefits of separating teaching and research careers for university faculty.

Using a standard model of incentives where agents allocate effort between two different tasks and ability is multidimensional, I show that the effect of stronger research incentives on research, teaching and sorting of professors depends on two main parameters: on whether teaching and research are substitute or complement in the professors' cost function and on whether teaching and research skills are positively or negatively correlated (i.e. whether good researchers are also good teachers). I then estimates the sign of these parameters. I overcome many of the standard identification issues by studying the case of Bocconi University, an Italian private institution of tertiary education based in Milan. Its institutional setting provides a unique opportunity to test the effect of research-oriented incentives on teachers' performance in multiple activities and the overall effect on the university teaching and research outcomes.

Three features of Bocconi's institutional setting are crucial for my analysis. First, I can construct a measure of teaching performance using a value added approach, that is the standard one used to evaluate teachers in primary or secondary schools (Rothstein, 2010; Rockoff, 2004; Aaronson et al., 2007; Rivkin et al., 2005). It is usually impossible to apply this method to universities because students self-select into courses, exams and teachers. Therefore, the usual assumption that - conditional on previous test scores - allocation is random, that may be credible in primary and secondary schools, does not hold in the university context.[1] Bocconi students are instead randomly assigned to teachers in each academic year: within a degree program, if the number of enrolled students requires it, students are randomly split in different classes, each of which taught by a different lecturer, while the exam, the syllabus and the type of classrooms are identical for all students. Therefore, once I include an exam script fixed effect, I can use the average class grade of students taught by different professors teaching the same course as a proxy for teaching quality.[2] Second, Bocconi sharply changed its faculty's incentive regime in 2005, shifting the focus explicitly towards research, by strengthening research requirements for promotion decisions and by introducing monetary incentives based on quality and quantity of publications. Third, the large heterogeneity of Bocconi teaching contracts provides a natural control group: many teachers are not fully hired by Bocconi and act only as external teaching faculty. They have the same teaching responsibilities but are not subject to Bocconi changes in promotion strategy and research incentives.

This paper therefore estimates a difference in difference equation, evaluating teachers' performance before and after 2005 and using external teachers as control group. The robustness of the results is confirmed by using two alternative control groups: faculty members who became tenured just before 2005, and are therefore not exposed any more to career concerns but only to the change in monetary incentives and faculty members of another Italian university (Bologna), very similar to Bocconi in terms of quality and quantity of research.[3]

My main results are as follows. First, the new incentive regime improved both the quality and the quantity of published papers. After the change in the incentive scheme Bocconi faculty members started to publish, on average, more papers than before, by 9% of a standard deviation. Moreover the effect is mostly driven by young faculty members, whose career concerns are stronger since they are not tenured yet. This result is in line with the literature on piece rate incentives (Lazear, 2000; Bandiera et al., 2007). Second, the introduction of incentives towards research had a negative impact on teaching performance, as measured by time-varying teacher fixed effect. In particular teaching quality decreased by 7% of a standard

[1]The only papers that can apply the value added method in the university context analyze the case of Bocconi university (Braga et al., 2014, Forthcoming) or the case of the U.S. Air Force Academy Carrell and West (2010) where the insituional setting implies student-teacher randomization.

[2]In particular I estimate time varying teachers fixed effects, controlling for yearly shocks at the course level-such as shocks to the exam papers or to the syllabus. In principle, once I include the course fixed effects, I do not need to control for students' characteristics such as previous test scores, because of randomization of students across classes within the same course.

[3]This second strategy can only be applied to Research outcomes, because there is no information on teaching performance for the university of Bologna. I chose Bologna, because in terms on quality of research as evaluated by the Italian Institute of University Research Evaluation (ANVUR) it is the most similar to Bocconi University, in terms of dimension of the department and quality of the research outcome between 2004-2010. www.anvur.org/rapporto/files/Area13/VQR2004-2010_Area13_Tabelle.pdf

deviation under the new incentive regime. The effect is, again, mostly driven by young faculty members and more negative for students at the bottom of the ability distribution. Combining the two estimates on teaching and research I find that, overall, one extra publication reduces teaching quality by one third of a standard deviation. This suggests that teaching and research are substitutes, not complement in the teachers' cost functions, at least for the type of courses I am considering. Third, there is evidence of some positive sorting effects: the new incentive scheme induced low ability researcher to leave the university, thus increasing the average quality of research at the university. Forth, I document that teaching and research skills are positively correlated: if a university manages to attract/maintain good researchers, it will also attract/maintain good teachers. The overall effect on teaching quality is therefore ambiguous: on one side, since teaching and research effort are substitutes, teaching quality of incumbents decreases, on the other side the policy pushes away the worst researchers and, since research and teaching skills are positively correlated, also the worst teachers.

This paper fits into the literature that investigates behavioural responses to incentives, in particular in the context of multi-task agents. As mentioned before, there is little empirical evidence of the actual cost of not optimally designed incentive schemes for multi-task agents, mostly because of limitations in the data and of the difficulty in measuring performance in many tasks, for instance because it is not observable or it is difficult to disentangle the individual contribution to the final outcome. Few exceptions, that usually analyse the quantity-quality trade off for the same activity, are Dumont et al. (2008); Feng Lu (2012); Hong et al. (2013); Johnson and Reiley (forthcoming). In the education literature Jacob (2005); Fryer and Holden (2013) analyze the impact of accountability policies on test-specific skills and students' effort in high-stake versus low-stake exams.

My paper contributes to the incentive literature, first, by providing a well-identified estimate of how multiple tasks interact in the agents' cost function. While most of the existing papers look at the quality-quantity trade off of performing the same activity, I analyse the effect on the performance in two different activities, when it is not clear a priori whether the tasks are substitute or complement in the agents' cost function. Second, to my knowledge this is the first paper that combines estimates of the effort substitution effect with an analysis of how multi-task agents sort in different types of firms, depending on their incentive schemes. This is key in order to evaluate the overall effect for the principal of different incentive schemes. Sorting effects may be very relevant and may countervail the direct effort substitution effect so to revert the sign of the impact of changes in the incentive scheme for all activities. Third, I am able to disentangle the pure effect of monetary incentives from the effect generated also by career concerns: this is extremely useful in order to understand the main drivers behind

different responses and to be able to efficiently design incentives in other contexts or settings. My paper is also related to the education literature on teachers' contracts and incentives. Some papers evaluate the effect of teaching contracts on teaching performances (Figlio et al., 2015; Bettinger and Terry, 2010; Figlio and Kenny, 2007) and find that students learn more from non-tenure line professors. Since non-tenure line faculty is less focussed on research, this may suggest that these results are driven by differences in teachers' incentive schemes. Still, it is impossible from these analyses to disentangle whether the effect they find is actually driven, instead, by selection into non-tenure line jobs. Some other papers look more directly at the trade-off between teaching and research, by analyzing the effect of increased teaching incentives on research and teaching outcomes. Brickley and Zimmerman (2001) use a single difference strategy to study the consequences of the introduction of teaching performance incentives at the University of Rochester Business School. The authors find a substantial and almost immediate jump in teaching ratings, measured by students' evaluations, and a corresponding decline in research output. Payne and Roberts (2010) analyze this same issue but using between, not within, university variation. They exploit US state variation in the adoption of teaching performance measures and find that research activity decreased in quantity but improved in quality in non-flagship universities.
This paper contributes to the education literature in two ways. First, it is the only one, to my knowledge, to test the other side of the relationship between teaching and research: the effect of strong research incentives. This type of analysis is crucial given the extremely wide adoption of research incentives in universities. Moreover, it is likely that the extent of effort reallocation generated by research incentives is larger than for teaching incentives because teaching effort is more difficult to measure and monitor and peer pressure on excellence in teaching is much weaker than in research. Second, this paper provides the first piece of evidence about the sign of the correlation between teaching and research skill. The positive correlation between teaching and research skills has important implications for the design of professors' incentives and hiring schemes. For example, policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that there is substantial overlap between good researchers and good teachers.

The structure of the paper is as follows. Section 2.2 provides a simple conceptual framework that rationalize expected results; Section 2.3 describes the data; Section 2.4 outlines the identification strategy; Section 2.5 presents my empirical results and Section 2.6 shows how my results are robust to alternative control groups. Finally, Section 2.7 briefly characterizes the policy implications of my results and concludes.

2.2 Conceptual framework

This section presents a very simple framework with the aim of organizing and rationalizing expected findings. The working of the model in the spirit of Holmstrom and Milgrom (1991) and it is similar to the model presented by Fryer and Holden (2013).[4]

An agent, upon accepting the contract, takes two non-verifiable actions e_r and e_t, which I call research and teaching effort respectively. Each action takes values in $\mathbb{R}_+$ and generates a performance measure $m_i = \alpha_i e_i$ where $i = r, t$ and α_i is unknown to the principal. I refer at α_i as the type of the agent on task i (her ability level).

I assume that the principal offers a linear incentive scheme of the form $x = s + b_r m_r + b_t m_t$. If the agent accepts, she makes her effort choices, the performance measure is realized and the principal pays the agent accordingly.

I also assume that the agent's preferences can be represented by the following CARA utility function:

$$u(x, e) = -exp[-\eta(x - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t)] \quad (1)$$

where x is the monetary payment and δ is the degree of substitutability between the tasks r and t in the cost function. Let $\underline{U}$ be the agent's outside option if he does not work. Moreover, I assume that there is a minimum teaching performance $\underline{m}_t$ and research performance $\underline{m}_r$ required by the university.

The agent therefore maximizes utility with respect to e_r and e_t, subject to the participation constraints ($u(x, e) > \underline{U}$ and $m_r > \underline{m}_r$ and $m_t > \underline{m}_t$). Note that when $m_r^* < \underline{m}_r$ or $m_t^* < \underline{m}_t$ each individual will choose whether to stay and exert effort level $\underline{e}$ (such that $\underline{m}_r = \underline{e}_r \alpha_r$)or to leave, depending on whether $U(x_{\underline{m}}, e_{\underline{m}})$ is larger or smaller than $\underline{U}$.[5] If it is smaller, she will decide to leave (or be fired). Otherwise, she will be induced to exert more effort, even if very costly, in order to stay in the university.

The new incentive scheme, that took place at Bocconi in 2005 as I will describe with more details in Section 2.4, implied an increase in b_r, the monetary return to research activity, and

[4] I will not model why the university decided to increase research incentives, i.e. I do not make assumptions on the university objective function, I only analyze what are the agents' responses to an increase in research incentive, in the spirit of Lazear (2000).

[5] $e_{\underline{m}}$ and $U(x_{\underline{m}}, e_{\underline{m}})$ are respectively the effort an agent needs to exert in order to obtain $\underline{m}$ and the utility level when $\underline{m}_r$ and or $\underline{m}_t$ are binding.

in $\underline{m}_r$, the minimum research performance required, but only for professors not tenured yet. Changes in b_r act mostly on the intensive margin (the amount of research effort to exert), changes in $\underline{m}_r$ instead mostly affect decisions also on the extensive margin (whether to stay in the university or to leave).

2.2.1 Effects on teaching and research performances

This section shows what happens to m_r^* and m_t^* (and therefore e_r^* and e_t^*) if the university increases b_r and $\underline{m}_r$ (to $\underline{m}_r'$) and professors remain in the university.

In the Appendix I solve the model (for internal solutions) and I show that the equilibrium effort level is:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{2}$$

It is clear that e_r^* increases if b_r increases, while the sign of the derivative of e_t^* with respect to b_r depends on the sign of δ.

Proposition 1. *An increase of b_r, the marginal return on research performance, leads to an increase in e_r.*

The response of e_t depends on the value of δ: $\begin{cases} \frac{\partial e_t}{\partial b_r} < 0 & if \delta > 0 \text{ (}e_r\text{ and }e_t\text{ substitute)} \\ \frac{\partial e_t}{\partial b_r} > 0 & if \delta > 0 \text{ (}e_r\text{ and }e_t\text{ complement)} \end{cases}$

The policy, moreover, increased $\underline{m}_r$.

Proposition 2. *When $m_r^* > \underline{m'}_r$: an increase in $\underline{m}_r$ does not have any effect.*
When $m_r^* < \underline{m'}_r$ and $U(x_{\underline{m'}_r}, e_{\underline{m'}_r}) > \underline{U}$, $\underline{m}_r$ is binding and professors exert $e_{\underline{m}_r,r}$ even if above their optimal level.

Where $m_r^* = \alpha_r e_r^*$; $U(x_{\underline{m'}_r}, e_{\underline{m'}_r})$ is the utility level achieved when research outcome $m_r = \underline{m'}_r$ and $\underline{U}$ is utility from leisure.

2.2.2 Sorting effects

Whether agents will decide to continue working under the new regime or to leave, depends on $\underline{U}$, the utility provided by leisure, $U(x_{\underline{m'}_r}, e_{\underline{m'}_r})$, the utility provided by achieving the minimum level of research performance in order to stay at the university and $U(x, e*)$, the utility provided by optimizing research and teaching efforts, without constraints.

Increases in b_r, do not have any effect on the decision to continue working or to leave the university because, at most, the agents will not change their behaviour. Increases in $\underline{m}_r$, instead, may have effects on the decision to leave the university.

Proposition 3. *If $\underline{U} > max \quad \{U(x_{\underline{m'}_r}, e_{\underline{m'}_r}); U(x, e*)|_{m*>\underline{m'}_r}\}$, professors will leave the university and enjoy utility $\underline{U}$*

Therefore, overall, for individuals whose $m_r^* > \underline{m}_r'$, the effect of the policy comes entirely from variations in b_r and therefore from evaluating the sign of the derivatives of e_r^* and e_t^* with respect to b_r.

For individuals whose $m_r^* < \underline{m}_r'$, the effect depends on whether $U(x_{\underline{m}}, e_{\underline{m}})$ under the new $\underline{m}_r'$ is larger or smaller than $\underline{U}$. If it is smaller, again, they will decide to leave and exert no effort. Otherwise, they will be induced to exert more research effort, even if very costly, and stay in the university.

I now evaluate how this effect varies by agent's ability. It is important to keep in mind that $\frac{\partial e_r}{\partial \alpha_r}|_{m_r=\overline{m}} < 0$: research effort is more costly for low α_r individuals. An increase in research incentives, therefore will be much more beneficial for high ability researchers. Instead, those more likely to leave because of an increase in $\underline{m}_r$ are low ability researchers.

Proposition 4. *When $m_r^* > \underline{m}_r$: an increase in b_r, leads to a larger increase in e_r for individuals with high α_r and to a larger response of e_t for individuals with low α_r.*
For low α_r agents, it is more likely that $\underline{m}_r'$ is binding, and that they are induced to leave.

The predicted response of stronger b_r along the distribution of α_r is therefore that: (i) for teaching, the effort substitution effect is stronger for low ability researchers (as long as $\delta > 0$); (ii) for research the effect is instead non-linear. Very low ability researchers will leave the university; of those staying, the lowest ability ones (those whose $m_r^* < \underline{m}_r'$) will increase effort on research in order to reach $\underline{m}_r'$; the others (those whose $m_r^* > \underline{m}_r'$), will increase e_r proportionally with their ability α_r.

2.3 Data and descriptive statistics

2.3.1 Students

This paper uses the administrative records of individual students and teachers from Bocconi University, an Italian private institution of tertiary education based in Milan. Bocconi offers degree programs in Economics, Management and Law. I only consider compulsory undergraduate courses between 2001 and 2011. My sample includes around 700 teachers and 30,000 students, who take on average 20 compulsory exams over the 3 years of study.

My data cover in detail the entire academic history of students, including their basic demographics (gender, place of residence and place of birth), high school leaving grades as well as high school type (whether the high school focusses on humanities, on sciences or technical/vocational subjects). Information is also provided on the grades in each single exam together with the date when the exams were sat. Moreover, I have access to the random class identifiers of students, which allow me to determine in which class each student attended each course.[6]

Table 1 reports descriptive statistics for students. Most of the students are Italian, one fourth is from Milan. They are positively selected among the population of high school graduates: the average high school final grade is very high (0.9 out of a maximum of 1[7]). On average there are 5 classes per course, of about 110 students each, and 20 compulsory undergraduate courses per year. Each student sits on average 7 exams per year. The degree program in Management is the one with the highest number of classes (7 on average).

2.3.2 Teachers

Together with data on students, I have access to administrative data on Bocconi faculty. In particular, I have information on teachers' demographics (date of birth, gender, full name),

[6]Students students who did not sit the exam in the academic year they were supposed to, are randomly allocated to a new class and the records on the initial class allocation are overwritten in the administrative database. I therefore include them in the new class, including a dummy equal to one if the student took the exam in a different year from what expected. However, this is a very small group (about 3% of students).

[7]Given that I know the maximum final high school grade each foreign student can take, I standardize high school final grades of foreign students to be between 0.6 and 1, so that they are comparable with grades of Italian students.

type of contract, department of affiliation and number of teaching hours in each course and class. I am therefore able to match students with teachers.

I classify each teacher as internal or external. Table 2 lists all different teaching contracts available at Bocconi over the years I consider and the way I group them into five categories: assistant professors-junior researchers, associate professors, full professors, non academics and professors from other universities. I define teachers in the first three categories as internal, treated by incentives, and teachers in the last two categories as external, my control group. I ecxlude lecturers, see Section 2.4.2

Table 3 reports descriptive statistics for teachers. Column (1) reports descriptives for internal teachers, column (2) for external teachers and columns (3) reports the difference of the two groups. In total, in my sample, I observe 681 teachers for 5 years on average. Internal teachers tend to be slightly older and to teach more hours at Bocconi. Most teachers are hired by the Management or Economics department. Finally, based on the data from 2005, one year before the change in incentives, internal professors represent about 70% of the sample.

2.3.3 Students-teachers randomization

The randomization of students to teachers is performed every year via a simple random algorithm that assigns a class identifier to each student, within each degree program.[8] Table 4 provides evidence that teachers were actually randomized to students. Following Braga et al. (2014, Forthcoming). I show results of a regression of class (student) average characteristics on teachers' characteristics and dummies for the full interaction of courses, academic years and degree programs.[9] The null hypothesis under consideration is the joint significance of the coefficients on teacher characteristics. The F statistics are always very low, suggesting there is no significant correlation between students' and teachers' characteristics.

[8] The university administration adopted the policy of repeating the randomization for each course with the explicit purpose of encouraging wide interactions among the students.

[9] This is the level at which randomization takes place

2.3.4 Publications

I collect publication data from the Web of Science website. In particular, I count professors' yearly publications in the fields categorized by Web of Science as 'business', 'maths' and 'economics'. Unfortunately, for less recent years, the Web of Science database only reports the author's first name initial and not the full name. As such, I run a search only using the authors' first name initial, together with their surname.

I also use Google Scholar as a source for the number of working papers. In particular, I use a web scraping program which makes automatic searches (one for each year/professor combination) from the Google Scholar website. I restrict my research on the Google Scholar website to the following fields: 'social sciences, arts, and humanities' and 'business, administration, finance, and economics'. In this case, data on full names are available for all years. I thus look for full names.[10]

2.4 Empirical strategy

This section develops my empirical strategy, aimed at estimating the causal effect of increasing incentives towards publishing on teaching and research performance.

I use administrative data from Bocconi university archives to estimate two Difference-in-Difference models, one for teaching and one for research, exploiting the sharp change in Bocconi research incentives and using external faculty as control group.

I begin this Section by describing in more details the reform in Bocconi's incentives regime announced in July 2005 (Subsection 2.4.1). Subsections 2.4.2 and 2.4.3 present my empirical model for the evaluation of the effect on teaching performance and on research activity respectively. Finally, Subsection 2.4.4 describes how I estimate sorting effects.

[10]This procedure does not eliminate the possibility that the same working paper is counted more than once, if published in two different versions. However, this is still a measure of the effort one puts in that specific research. Moreover, this measure also contains the published version of the working papers. Accessed in December 2011.

2.4.1 The new incentive policy

In 2005, Bocconi University unexpectedly announced the adoption of a new policy of hirings and promotions. The Board of Directors called for the Rector to make Bocconi University one of the top five universities in Europe. As a consequence, the old hiring and promotion strategies, mainly based on national competitions and seniority, were replaced with new practices based on international standards. Since then[11], an independent committee, composed of faculty members from all disciplines, has been in charge of recruiting and promotions. Decisions have been centralized at the university level, making exceptions impossible. Moreover, the importance of research outcomes in promotion decisions was clearly stated in all internal faculty contracts.

The goals of the New Strategic Plan, as announced in July 2005, were the following: (i) recruiting at least 50% of new faculty on the international job market; (ii) improving the systems to evaluate research produced by each professor (through the creation of an independent evaluation committee and the internationalization of evaluation criteria); (iii) adopting clear incentives on research (both monetary[12] and career-based); (iv) creating mechanisms to "attract and keep the best researchers worldwide".

The focus switched explicitly towards research, tenure decisions started to be based almost entirely on scientific productivity and the requirements on quantity and quality of research started to be much tighter.

2.4.2 Research performance

I first evaluate whether incentives on publishing have an impact on research quality or quantity.

I use three different measures of research performance: (i) the number of publications as

[11]The actual implementation of the policy was in 2007, but throughout the analysis I will consider the year of the announcement, 2005, as the treatment year. Be aware that the full effect will be in place starting from 2007.

[12]Even if previously anticipated, Bocconi started to actually provide monetary incentives to its internal faculty in the academic year 2008. In particular there are three types of incentives: (i) the possibility of getting "research profile", with less teaching duties; (ii) research premia that depend on the number and the quality of publications; (iii) research funds, given to everybody who has reached a minimum level of research productivity in the previous two years. Publications were weighted depending on the quality of the journal)

collected from Web of Science; (ii) a proxy of the index actually used by Bocconi to evaluate teachers (which is computed as the sum of the number of articles published by each teacher, weighted by the quality of the journals as classified by Bocconi[13], divided by the number of co-authors) and (iii) the number of working papers and published papers (from Google Scholar).

I then implement a Difference-in-Difference model by estimating the following equation for the years between 2001 and 2010:

$$pub_{pt} = \theta_t + \theta_p + \gamma_{res}(internal_p * post2006_t) + \gamma_4 X_{pt} + \eta_{pt} \tag{3}$$

where pub_{pt} are publications of professor p in year t; $internal_p$ is the internal status (in 2005); θ_t are time fixed effects; θ_p are teacher fixed effects; X_{pt} are teacher characteristics (age, age squared) and η_{pt} is the error term. I cluster standard errors by professor.

For sake of consistency, I include only teachers who were teaching classes I can use to estimate the teaching equation (see below equation 5).[14] Moreover, in order to exclude endogenous status switches from internal to external or vice versa after the introduction of the policy, I classify teachers as internal if they were internal in 2005, before the change in promotion strategy. In my robustness checks (Section 2.5.3) I check my results are not driven by this choice, by running the same analysis using contemporaneous status instead of status before 2006 as treatment, therefore including endogenous 'switches' in the effect. Moreover, I drop internal lecturers. Lecturers are internal professors (fully hired by Bocconi) but with only teaching duties.[15] On one side, monetary research incentives are not provided to lecturers but, on the other side, the way lectureship decisions are taken has probably changed after 2006. They therefore do not represent a good control group. In a robustness check (Section 2.5.3), I include lecturers and interact them with the treatment. Finally, in order to use the same sample of teachers as in the analysis on teaching, I do not consider law professors and law courses: law's exams are usually oral exams so the set of questions is not the same for all students. It is therefore difficult to use average grade as a measure of teaching quality.

[13] Bocconi divides journals into 3 categories: A+ journals (i.e. Econometrica), to which it assigns a weight of 15; A journals (i.e. Economic Journal), weighted 7; B journals (i.e. Economic Letters), weighted 3. I classified journals using the list valid for the year 2007, available upon request.

[14] The difference in the number of observations is given by those teachers who were teaching more than one class per year or by the fact that some teachers do not teach compulsory undergraduate courses all years, but I still include those year observations in my analysis, for consistency over time.

[15] The difference between the position of lecturers and assistant/full professors is clear from how their contracts. The contract for assistant professors states "responsibilities include teaching and, most importantly, productivity in research". The contract for lecturers, instead, states that only teaching duties are expected from lecturers. Research activity is not even mentioned.

2.4.3 Teaching performance

Second, I estimate my empirical model for the effect on teaching. I do it in two steps.

The Bocconi institutional setting allows me use average students grades, controlling for yearly shocks at the exam paper, as a proxy for teaching quality. Students taking the same course are all taught the same syllabus and are all examined on the same questions, independently of the class to which they are (randomly) assigned. Some variations in the material and in the exam across degree programs are allowed (this is why I correct for the full interaction of courses, degree programs and years). Usually a senior member of the faculty acts as the course coordinator: he establishes the material to teach, manages possible complications and prepares the exam paper. Grading is instead generally delegated to the individual teachers, who typically are supported in the marking by teaching assistants.

The first step of my teaching analysis uses micro data from the student academic curriculum database and it is aimed at computing the average grade at the class level, conditional on students' high school final score and demographics.[16] I estimate the following equation:

$$grade_{ipct} = \beta_0 + \beta_1 HSgrade_i + \beta_2 X_i + \alpha_{ptc} + u_{ipct} \tag{4}$$

where $grade_{ipct}$ is the grade obtained by student i, with teacher p[17], in year t, in course c (standardized at the course-year level to have mean 0 and standard deviation 1); $HSgrade_i$ is student i high school final grade; X_i are the students' individual characteristics (gender, age, whether Italian, whether from Milan, type of high school attended); u_{ipct} is the error term. α_{pct}, the year specific teacher fixed effect, is my parameter of interest.

The second step evaluates how the time-varying teacher fixed effects α_{ptc} evolve over time, in response to the change in incentive regime. I implement the same Difference-in-Difference estimation as in Section 2.4.2, changing the dependent variable. In particular, I estimate the following equation:

$$\widehat{\alpha_{ptc}} = \delta_p + \delta_{tc} + \gamma_{teach}(internal_p * post2006_t) + \gamma_2 X_{pt} + \epsilon_{ptc} \tag{5}$$

[16] To reduce computational burden, I exploit randomization of students to teachers and I do not include students fixed effects.

[17] Since in around 40% of the cases more than 1 professor teaches the same class the actual meaning of p in this first case is the "professor mix" of the class.

where $internal_p$ is a dummy equal to one if the professor was internal before the change in incentives; $post2006_t$ is a post reform dummy; δ_p are teacher fixed effects[18]; δ_{tc} are fixed effects for the full interaction between academic years, courses and degree programs[19]; X_{pt} are time-varying professor characteristics (age, age squared, experience in teaching undergraduate courses in Bocconi) and ϵ_{ptc} is the error term. I cluster standard errors by professor.
γ_{teach} quantifies the change in teaching performance of incumbent professors under the new incentive scheme more focussed towards research.

The economics literature usually measures teacher quality by estimating a teacher fixed effect in equation 4. Here I allow teacher effects to vary over time and I analyze how they change in response to the positive shock in research activity.
I overcome many of the standard identification problems because: (i) I eliminate concerns related to time constant factors by including teachers' fixed effects in my regressions: I only analyse how teaching performance evolves over time; (ii) I eliminate concerns related to time varying endogenous matching thanks to the randomization of students to teachers. There has been a debate (Rothstein (2010, 2009); Ishii and Rivkin (2009); Kane and Staiger (2008); Chetty et al. (2014)) about whether value-added models perform weakly in the absence of randomization. Teachers fixed effects may also identify endogenous matching between teachers and students. Results are mixed. Most recently Kane and Staiger (2008); Chetty et al. (2014) use primary school data to show that this problem can be eliminated controlling for previous year test scores. However, endogenous matching is likely to be much stronger in the university context, where students self-select into courses and therefore teachers.

Finally, I also estimate the same effect running the analysis directly at the student level, as follows:

$$grade_{iptc} = \zeta_p + \zeta_{tc} + \zeta_{teach}(internal_p * post2006_t) + \zeta_2 X_{ipt} + v_{iptc} \tag{6}$$

where all the variables are defined as before and v_{iptc} is the error term.
While my preferred specification is the estimation of equation 5, because it is more easily interpretable in terms of changes in teaching quality at the professor level, this last specification will allow me to evaluate how the main effect is heterogeneous with respect to students' characteristics, in particular with respect to students' previous test scores, measured by their final high school grade.

[18] Notice that in this case p represents a single teacher. Therefore if a class was taught by multiple teachers I impute the (unique) class fixed effect to both teachers.

[19] Courses may have the same code but programs and exams may be different for different degree programs. Interacting also with degree programs allows me to exploit variation across teachers' performance when syllabus and exam papers are exactly the same (and over which the randomization of students to teachers takes place).

2.4.4 Sorting patterns

To have a complete picture of the overall effect of the change in incentives on research and teaching quality, I analyse how the composition of workers changed after the new regime was introduced. As shown in Section 2.2, the change in minimum research requirements should push low ability researchers away. Whether this translates into maintaining also better teachers, it will depend on how teaching and research skills are correlated.

I analyse selection effects in two ways: first, I compare estimates with and without professors' fixed effects; second, I obtain direct estimates of the underlying teaching and research abilities and I analyse how the ability composition of teachers varies over time, looking both at teachers sorting in and sorting out.

In order to analyse sorting patterns, I need estimates of teaching and research skills. I obtain these estimates using professors' fixed effects obtained from the following equations.
For teaching:

$$\widehat{\alpha_{ptc}} = \theta_p^t + \delta_{tc} + \gamma_2 Q_{pt} + \epsilon_{ptc} \tag{7}$$

where α_{ptc} is the conditional average grade of professor p, teaching course c in year t; δ_{tc} are fixed effects for every course-degree program-year combination; Q_{pt} are professors' characteristics (age, age squared, years of experience at Bocconi); ϵ_{ptc} is an error term. Finally, θ_p^t are professor fixed effects, my estimate of underlying teaching ability.

Analogously, for research:

$$pub_{pt} = \theta_p^r + \zeta_t + \zeta_2 Q_{pt} + \eta_{pt} \tag{8}$$

where pub_{pt} is the number of papers published by professor p in year t; ζ_t are year fixed effects, that absorb any possible time trend in how difficult it is to publish papers over time; Q_{pt} are professor characteristics (age, age squared and their department of affiliation); η_{pt} is an error term. Again, θ_p^r are professor fixed effects, my estimate of underlying research ability.

One first concern is that, since incentives are muted under the new scheme, it is not clear whether fixed effects based on teaching or research productivity after 2006 are a good proxy for ability. This would imply one should only use fixed effects evaluated before 2006. However, it would be impossible to test whether the new policy managed to attract high ability professors, since I would not be able to estimate a teacher fixed effect for faculty members who entered under the new incentive regime. In Figure 4, I follow Lazear (2000) and I show that, for professors who were teaching also before 2006, there is a strong positive correlation between fixed effects evaluated in the period before 2006 and those estimated for the period after 2006. Whenever it is possible (for sorting out effects), I will run my regressions also using fixed effects estimated on the pre-2006 period only.

For this analysis I do not run a proper difference-in-difference strategy because I do not know the entire employment history of external teachers. I only observe whether they were teaching undergraduate compulsory courses between 2001 and 2011, but I do not observe their exact year of entry/exit. For internal professors, instead, I know exactly their year of entry, every change in their contracts and their year of exit, including the reason for leaving. Moreover, it is very unlikely that external teachers represent a good control group for the analysis on sorting: the way they are selected is very different from the selection process of internal faculty and it varies substantially depending on specific departments and academic years.

I evaluate how average teaching and research ability change, depending on the year teachers entered/exited Bocconi.
For sorting out, I estimate the following equation:

$$\widehat{\theta_p^j} = \alpha_1^j exitpost2006_p + \alpha_2^j exitpre2006_p + \alpha_3^j X_p + \delta_e^j + u_p^j \tag{9}$$

where: $j = r, t$ refers to research and teaching, respectively; δ_e^j are year of entry fixed effects; $exitpost2006_p$ is a dummy equal to one if teacher p left Bocconi after 2005; $exitpre2006_p$ is a dummy equal one if professor p left Bocconi before 2005[20]; X_p are time-invariant professors' characteristics (age of entry, gender) and u_p^j is an error term. I only include teachers leaving Bocconi for reasons different from retirement.
Symmetrically, I obtain the effects on sorting in of teachers, by estimating the following equation:

$$\widehat{\theta_p^j} = \psi_1^j entrypost2006_p + \psi_3^j X_p + \psi_4^j f(e) + \omega_p^j \tag{10}$$

where: $j = r, t$ refers to research and teaching, respectively; $f(e)$ is a linear and squared trend for year of entry; $entrypost2006_p$ is a dummy equal to one if teacher p entered after 2005; X_p are time-invariant professors' characteristics (age of entry, gender) and ω_p^j is an error term. To make the two groups of teachers more comparable, I estimate equation 10 only for teachers who entered after 2000.

[20]the omitted category are those staying

2.5 Results

2.5.1 Results for research

The sign of the effect on research is expected, from Section 2.2, to be positive and stronger for young professors not tenured yet, since they are affected both by the changes in monetary incentives and by the changes in promotion strategies.

Table 5 shows some descriptive statistics for the number and the quality of publications and working papers for internal and external teachers before and after 2006. The first panel analyses the total number of publications (books or journal articles) of professor p in year t, as collected from the Web of Science database. The second panel looks at the number of publications, weighted by the importance they have in terms of Bocconi's new incentive regime. This allows me to evaluate quality as well as quantity of research. The first column reports the mean and the standard deviation of publications for internal and external teachers. The second and the third columns break down the number of publications for the period before and after 2006. Finally, the number in the bottom-right corner represents the simple difference in difference, without any control. Standard errors, clustered at the teacher level, are reported in parenthesis. The Table shows that the number and the quality of publications increased after 2006 and they increased much more for internal professors than for external professors.

Table 6 shows results from equation 3, using the three dependent variables described above. Columns (1) and (2) report estimates without teacher fixed effects. The effect is positive and significant in all three panels. Once I include teacher fixed effects (columns (3) and (4)) the effect is still positive and significant. After the introduction of research incentives, the number of publications increased by 0.14 (9% of a standard deviation) for internal faculty and the index used by Bocconi to evaluate teachers increased by 0.13 (6% of a standard deviation). The number of working papers of internal professors is also 0.15 (6% of a standard deviation) higher than it would have been otherwise. Moreover, while columns (1) and (3) look at the aggregate effect, columns (2) and (4) separately evaluate the effect for for assistant professors and associate professors (which I call junior faculty) and full professors. The aggregate effect is mostly driven by junior faculty, as their career concerns are stronger.
Finally columns (5) and (6) report results from estimating equation 3, using as dependent variable the square root of the number of publication. This is to try tackle simultaneously

the presence of possible outliers and of a lot of zeros.[21]

Figure 1 displays the evolution of the difference in average number of publications between internal and external faculty.[22] The dotted lines refer to the 10% confidence interval boundaries. While the difference is rather stable before 2005, it gets larger after the introduction of research incentives. Moreover, given the long time needed to publish papers in most disciplines, after 2006 there is a clear change in trends but there is not a sharp jump.

2.5.2 Results for teaching

As shown in Section 2.2, the sign of the effect of stronger research incentives on teaching quality depends on whether teaching and research efforts are complements or substitutes in the professors' cost function (δ smaller or larger than 0 respectively). The effect moreover is expected to be stronger for junior professors, exposed both to the change in monetary incentives and to the change in the minimum number of publications required.

Table 7 presents the results obtained from estimating equation 4. Exam grades are standardized to have mean 0 and standard deviation 1 within the same course-year.
Results show that being male, with a higher final high school grade, Italian and from Milan is associated with higher university exam grades.[23]

Table 8 reports some summary statistics of the estimated α_{pct} for internal and external teachers, before and after 2006. While before 2006, the teaching performance of the two groups was very similar, after 2006 it improved much more for external teachers than for internal teachers. Again, the bottom-right corner reports the difference in difference, without any control.[24]

Table 9 displays results from estimating equation 5. The first two columns show results without teacher fixed effects. Column (3) and (4) add teachers fixed effects. Teaching quality

[21] Moreover I dropped the 5/1000 highest values for each dependent variable, since it is very likely that most outliers are generated by homonymity.

[22] This graph plots the coefficient γ_s of the following equation (where $\theta_s = 1$ if $t = s$):

$$pub_{pt} = \theta_t + \theta_p + \Sigma_{s=2001}^{2011} \gamma_s (internal_p * \theta_s) + \gamma_4 Q_{pt} + \eta_{pt}$$

.

[23] Grades in Italy go from 18 (pass) to 31 (excellence).

[24] Notice that, because of some sampling error generated by the fact that $\hat{\alpha_{pct}}$ are estimated, the reported standard deviation may be larger than the standard deviation of the true α_{pct}.

of internal teachers is 0.04 (around 7% of a standard deviation) lower after the change in incentives than it would have been otherwise. This suggests that teaching and research are substitutes in the professors' cost function. Again, the effect is stronger for young faculty members, more exposed to the policy.

Panel a and b of Figure 2 show the evolution of the different performance of external and internal teachers (panel a) and external and assistant professors (panel b) over time[25]. The difference is rather stable before the academic year 2005/2006 (named 2006 in the graph). Right after the adoption of the new incentive regime there is a drop in the quality of teaching for internal professors. In the following years, the performance is still slightly worse than before the reform, but better than in 2006. This may be because internal professors understood the consequences of their effort reallocation and partially readjusted their behaviour. Alternatively, they just started being more generous with their grading standards. Section 2.5.3 analyzes this aspect in more details.

Table 10 reports results from the student level regression (equation 6).[26] As expected, results are very similar. What differentiates columns (1) and (2) of Table 10 from columns (3) and (4) of Table 9 is the way observations are weighted and coefficients should be interpreted. Table 10 implicitly weights observations by the number of students in each class: the coefficients should be interpreted as effects on average students' performance. Table 9 weights observations by teachers and the coefficients should be interpreted as effects of average teachers' performance.

Columns (3) and (4) of Table 10 explore whether the main results of Table 9 mask some important heterogeneity at the student level. I estimate equation 10, interacting the main effect with a proxy for students' ability. In particular I use high school final grade as proxy.[27] My omitted category are high ability students. Results show that the negative effect is mostly borne by low ability students.

This result suggests that there is room for policies aimed at matching professors to students in order to reduce the overall negative effect of stronger research incentives on teaching performance. This would mean in this case to match young researchers, more affected by the

[25] This is obtained by plotting the coefficients γ_s obtained from the following equation (where $\delta_s = 1$ if $t = s$):

$$\alpha_{ptc} = \delta_p + \delta_{tc} + \Sigma_{s=2001}^{2011} \gamma_s (internal_p * \delta_s) + \gamma_2 Q_{pt} + \gamma_3 Z_{pct} + \epsilon_{ptc}$$

Year 2001 (and the interaction between 2001 and internal) is omitted. The dotted lines refer to the 10% confidence interval bands.

[26] In this case, whenever a class was taught by more than one teacher, the observations for each student were doubled, such that each student was imputed to every teacher he was assigned to.

[27] I divide it into 3 categories: (i) high ability (omitted)= those students whose final high school grade was between 1 and 0.9; middle ability = between 0.8 and 0.9 and low ability: below 0.8.

change in incentives, to higher ability students, who are less damaged by their lower teaching quality.

2.5.3 Robustness checks for the effect on teaching

Table 11 presents a set robustness checks for the estimation of the teaching equation. First, I estimate equation 5 excluding the academic years 2008/2009, 2009/2010 and 2010/2011. Starting from 2008/2009, internal faculty was exposed not only to research incentives, but also to teaching performance monetary awards. In particular, Bocconi University created a commission in charge of awarding a premium of 20,000 euros for the best 20 teachers who voluntarily apply. Decisions are based on students' evaluations. This new policy may attenuate the effect of research incentives (Holmstrom and Milgrom (1991, 1994)). Column (1) of Table 11 shows that results are almost unchanged. Second, in Column (2) I include lecturers in my sample and I estimate a different treatment effect for lecturers. The effect on internal professors is similar. The effect on lecturers, even if not significant because of the small number of observations, is negative. Column (3) includes endogenous switches from internal to external status after the policy: it uses the contemporaneous status, not the status before 2006 as in Table 9, to define internal status. The control group includes in this case also, for instance, professors who switched from internal to external as a consequence of the policy. The coefficient is still negative and significant, but the magnitude is smaller. This means that Bocconi promotions from external to internal and viceversa where positively correlated with teaching quality. In column (4) I weight my regression by the number of hours taught by each professor in each class. The results are again very similar.

I now discuss three possible confounding factors, that may undermine my identification strategy. The first is that students might not comply with the random class assignment and they might endogenously decide to attend classes with different lecturers. For example, they may match to the best professors, or attend classes with their closest friends. Unfortunately, I do not have any direct information on these unofficial switches of classes.[28] Braga et al. (2014) analyze whether the direction of class switches at Bocconi University is correlated with professors' ability. They use data on students' answer to an item in the student evaluation

[28] It would have been in principle possible to grasp the size of classes reallocation by using students' evaluations, exploiting the information for whether the number of answers is larger or smaller than the official class size. However, Bocconi decided to hand in evaluation forms to a sub sample of professors only exactly in the year 2005 and 2006, making it impossible to look at students' evaluations for the period when the policy took place.

forms asking them about the level of congestion in their classroom. They estimate the degree of class switches as the difference in congestion level between the most congested and the least congested classes for each course. They find that, overall, course switching is not related to teacher effectiveness in any direction. Therefore, if the process of class switching is unrelated to teachers or students quality, then it will just affect the precision of my estimated class effects. Moreover, if the process is constant over time, the effect will go away with professors' fixed effects. Finally, even if course switching does affect my results, it would probably bias them against finding a negative effect on teaching performance. It is likely that students, if anything, will react by attending classes with the best teachers, who after the change of incentives will more likely be external faculty members. This would reduce the negative effect of the incentive policy on teaching.

Another concern is that teachers may change the way they grade students' exams as an effect of observing worse performances of their students. In this case the observed change in teaching quality may actually be confused with a change in grading standards. There is not a common rule on how exams are graded at Bocconi: in some cases exam papers are randomly allocated to be graded to one of the course teachers independently of the class they were assigned to, in some other cases each professor is in charge of grading his own group. I do not have information on how exam papers are actually graded in each course. However, in the fist place, if anything, I expect internal teachers to start being more lenient towards their students, therefore I expect this type of bias to go against finding a negative effect on teaching performance. Moreover Table 12 addresses this point by looking at subjects were grades are more difficult to be manipulated. Columns (1) and (2) look at the effect on teaching quality for exams that are more objectively-graded, such as math, statistics or quantitative finance. Results show that, even if the effect is slightly smaller and less precise for this types of courses, it remains negative.[29]

Another concern is that internal faculty may have managed to reduce its teaching loads and to avoid some of its previous teaching responsibilities, under the new regime. In columns (3) and (4) of Table 12, I check whether the new incentive scheme implied a change in internal teachers' teaching duties. I estimate equation 5 using as dependent variables a dummy equal to one if professor p was the course coordinator in year t and the number of teaching hours taught by professor p in year t, respectively. Results show that there is no significant change in the type of teaching loads and duties before and after the change in the incentive regime. This suggests that the change in teaching quality was not driven by other, simultaneously

[29]Notice that it may be that what generates these results is just the fact that teaching are research efforts are more complement for math subjects than for other subjects.

related, changes in how teaching was organized and distributed.

Finally, in Table 13, I check whether my results may be driven by a recommendation letter sent by Bocconi University in 2006 to the entire teaching faculty, asking for higher homogeneity of grades across classes. This may affect my analysis, if internal and external teachers responded to this request differently. Table 13 displays the standard deviation of average class grades across classes belonging to the same course and degree program, by academic year and by whether the teacher was internal or external. The variability of grades between classes did not decrease right after 2006, as a consequence of such recommendation, and there was no differential response between internal and external teachers.

2.5.4 Teaching and research skills

Understanding the sign of the correlation between α_r and α_t, as defined in Section 2.2, is crucial both to have a full picture of potential sorting effects and to understand the plausible cost of separating careers of teachers and researcher in university.

Figure 5 and Table 15 correlate the two sets of fixed effects as estimated from equations 7 and 8 and they show that teaching and research ability are strongly positively correlated: good researchers are also good teachers. This is an important result that has not been estimated before. Columns (1) and (3) include all teachers in my sample. Columns (2) and (4) try to address the fact that teacher fixed effects represent noisy measures of the true teaching and research abilities and sampling error may bias the coefficients of columns (1) and (3). I exploit the fact that sampling error decreases substantially if the analysis is performed on a subsample of teachers with a large number of observations. I therefore estimate the correlation, including only teachers for which I can estimate the fixed effects with more that 5 observations.[30] Results are very similar but, as expected, after the correction the coefficients are larger, because not affected anymore by the attenuation bias.

The positive correlation between research and teaching skills and the large standard deviation of the teachers fixed effects have important policy implications. First, comparing the standard deviation of the fixed effects plotted in Figure 5, which quantify the variation in the time-invariant part of teaching quality (in teaching ability), with the coefficients obtained in

[30] Notice that I always estimate the research fixed effect with 10 (yearly) observations. For the teaching fixed effects, instead, the number of teacher-specific observations used depends on the number of time I observe teacher p teaching undergraduate compulsory courses.

Tables 6 and 9, it is clear that sorting effects may potentially have much larger and substantial consequences on the overall research and teaching productivity than substitution effects. Keeping the composition of teachers constant (i.e. looking at the intensive margin), the reform of the incentive structure improved research productivity by around 9% of a standard deviation and decreased teaching quality by 7% of a standard deviation. Instead, when we allow the composition of teachers to change (i.e. we look at the extensive margin) and we incorporate the fact that universities will, as a consequence, attract (push away) the very best (worst) researchers, the average productivity may potentially increase by much more. Therefore, even if Section 2.5.2 showed that, at the margin, pushing university professors to focus more on research may induce them to crowd out time from preparing teaching classes and may worsen their teaching performance, Figure 5 shows that sorting effects may potentially be much more effective.

Second, the fact that teaching and research ability are positively correlated entails that if universities are able to attract/maintain good researchers, they will also, indirectly, improve teaching quality. A very popular proposal to solve the trade-off between teaching and research, is to increase specialization of faculty members. This would entail, for example, the creation of two groups of professors, one more research-oriented and one more teaching-oriented. Figure 5 and Table 15 show that these proposals should take into consideration that good researchers are also good teachers and the potential benefit of separating careers may be minimal, given that the trade-off on the intensive margin (generated by forcing good researchers to teach some class instead of researching) is less sizable than the trade-off on the extensive margin (generated by excluding good researchers from teaching, for instance).

2.5.5 Sorting

The first way I analyse sorting effects is by evaluating the difference between the OLS and the fixed effect estimates in Tables 6 and 9. OLS estimates are always larger than fixed effects estimates, suggesting that the policy induced some positive sorting effects.

I also analyse sorting in and out separately using direct estimates of teachers' underlying ability, obtained through equations 7 and 8.

Table 16 shows how teachers' fixed effects change for (internal) professors fired before and after the change in the incentive regime, for research ability and teaching ability respectively. Columns (1) and (2) use fixed effects estimated for the entire period. The dependent variable of columns (3) and (4) are, instead, fixed effects estimated for the pre 2006 period only.

Results, in line with the predictions of Section 2.2, show that the change in incentives induced worse researchers and therefore worse teachers to leave.

Table 17 reports instead results from equation 10 and it shows no effects on Bocconi's ability of attracting good teachers or good researchers. The reason why I don't find any positive sorting-in effect, partly in contrast with what is expected from the results of Section 2.2, may be due to the fact that it takes time to publish papers and it may be too early to evaluate the research and teaching productivity of very young scholars.

2.5.6 Heterogeneity by teachers' ability

This Section analyses how the effect on teaching and research performances changes with respect to teachers' ability. Section 2.2 shows that the bulk of the effort reallocation should be concentrated on low ability researchers, while the effect on research should be concentrated on very low ability (because of fair of being fired) or very high ability (because they benefit more from any unit of effort in research) researchers.

Column 1 of Table 14 shows that the negative effect on teaching activity is stronger for low ability researchers than for higher ability ones. The Table displays the coefficients of the $internal_p * post2006_t$ dummy of equation 5 interacted with research ability tertiles.[31]

For what concerns the heterogeneity of the effect on research performance, columns 2 and 3 of Table 14 show that the positive effect is driven by low and middle ability researchers. The difference is more evident in column 3, that looks at the effect on the number of working papers.

2.6 Alternative control groups

One possible concern of using external teachers as control group is that these teachers may react to the policy as well if, for instance, their final objective is to be hired by Bocconi. This would spoil my identification strategy because it implies that the effect of the policy

[31] Tertiles are calculated using θ_p^r of equation 8, and are estimated only for the years before the change in the incentive regime. This is to avoid that the way ability is measured is affected by the change in incentive regime itself.

would spill over my control group. Moreover, one may think that external teachers are a natural control group for evaluating the effect on teaching performance but may not be as good as a control group for research activity, because they may have very different research productivity and may be on very different trends in any case. To tackle these issues I propose two alternative control groups.

The first one refers to the analysis on research. In Table 19, I use all professors belonging to Bologna University faculty in 2005 as alternative control group. Bologna University is another Italian University, whose department of management and economics is quite similar to Bocconi University in terms of quality of the economics/management department. Bologna university economics and management department is indeed ranked as the best[32] department among Italian public institutions. Table 18 shows the productivity of Bologna University faculty members in terms of research, compared to Bocconi's faculty members. Again, I obtain data on their publications from the Web of Science website and data on the faculty composition in 2005 from the website of the Italian Ministry of Education.[33]

The second alternative control group are professors who became tenured before the policy. Given that the change in the incentive structure acts mainly in terms of promotions and tenure decisions, full professors should only be marginally affected. Since they are already fully hired by Bocconi, they should not react to changes in hiring/promotion strategies. If we assume that the effect of monetary incentives is the same on full and junior professors, than what I estimate using full professors as control group is the effect of the change in career requirements only. However, for publications it is very likely that trends for junior and senior professors are different, after they get tenured, since tenure decisions are based research productivity or potential productivity. I will therefore use this alternative control group only for the analysis on teaching.

Moreover, both for the analysis on teaching and on research, I estimate my difference-in-difference models separately on two sub samples of professors with similar age. In particular I split both the sample of internal and external teachers between those older than 43 (the median age) and those younger than 41. This allows me to use as control group for young researchers, young external researchers since, especially for research, junior and senior faculty members may be on very different trends.

Table 19 reports results for research activity. In columns (1) and (2) I run equation 4 on the

[32]Or one of the top three departments, depending on the type of ranking.
[33]www.miur.it

subsamples of teachers younger and older than 43, respectively. As for the main results, the effect is larger for junior professors. Columns (3), (4) and (5) use, instead, Bologna faculty members as control group. Column (3) looks at the aggregate effect, column (4) compare junior professors at Bocconi with junior professors at Bologna and column (5) compares full professors. The effect is remarkably similar to my baseline estimates. The introduction of incentives led to an increase in the number of publication of 0.17 for Bocconi faculty members, very similar to the effect found in Table 6. The increase is stronger for young faculty members.

Table 20 reports, instead, results using alternative control groups for teaching quality. Columns (1) and (2) split again the sample by age. The effect is similar to what found in my baseline estimates and is more negative for junior professors. Columns (3) and (4) use full professors as alternative control group. Columns (3) does not include teacher fixed effects. Column (4) shows results including teachers fixed effects. Again, results are remarkably similar to what found in Table 9. The introduction of research incentives worsened teaching performance by 0.04, about 7% of a standard deviation. I can't use Bologna faculty members as control group for the analysis on teaching, because information on teaching performance of Bologna faculty members is not publicly available.

Figure 3 checks the presence of parallel trends.

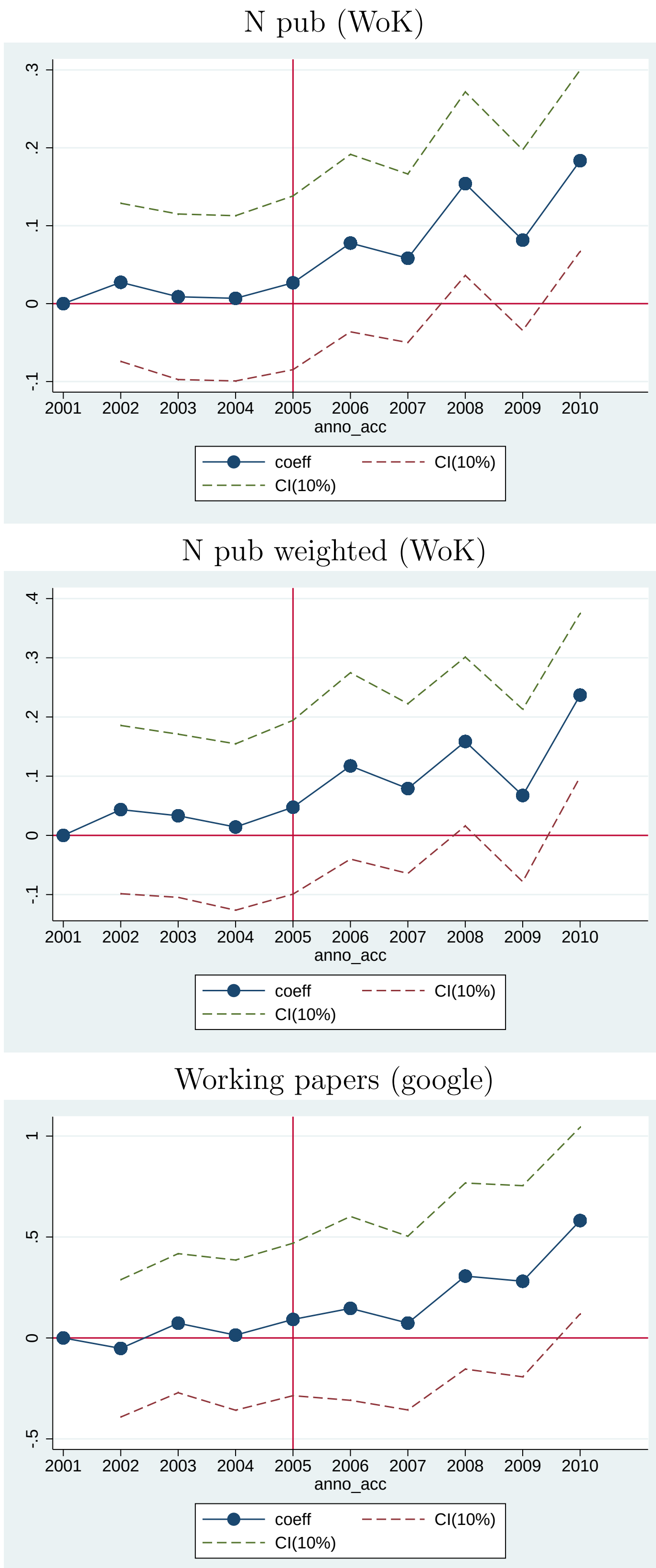
2.7 Conclusions

This paper exploits a natural experiment to test predictions of models of incentives in a multi-task environment. I use administrative data from Bocconi University to analyse faculty reaction to a sharp increase in research incentives. The heterogeneity in the teaching faculty type of contracts allows me to find a control group for my Difference-in-Difference estimation. The randomization of teachers to students within the same course, in a context where the syllabus and the exams are fixed, allows me to build a credible measure of teaching performance. In particular, the specific Bocconi setting allows me to overcome two of the reasons why analyses of teachers' effectiveness are rarely done at the post secondary level: the lack of standardized tests and the endogeneity in students selection of courses (and professors).

I find evidence that the introduction of research incentives affects the allocation of effort across tasks. Results show that professors' teaching performance gets worse while their research performance significantly improves. In line with the predictions of Holmstrom and Milgrom (1991, 1994), I find that the effect is stronger for young faculty members, more exposed to career concerns. This provides evidence of the importance of implicit and explicit incentives in an organization. The number of working papers and published papers of internal Bocconi faculty increases after the introduction of incentives on research. The magnitude is in line with the literature on provision of incentives (see Prendergast (1999); Lazear (2000); Checchi et al. (2014), for example). I observe that the effect on quantity of publications does not go against the quality of publications. This may be due to the way research incentives are structured by Bocconi. On the other hand, teaching quality of faculty members more exposed to research incentives is 7% of a standard deviation lower after the change in the incentive regime. The effect is nonproportionally borne by lower ability students. My estimates suggest that encouraging one more paper has an implicit cost of 0.3 standard deviation on teaching quality. Moreover, I find evidence of positive sorting effects. After the change in incentives, lower quality researchers left Bocconi faculty and, since teaching and research ability are positively correlated, the policy attracted also good teachers.

My results suggest that it is beneficial to evaluate new policies not in isolation but as part of a coherent incentive system. I believe this paper delivers three important policy-relevant messages. First, since the negative effect on teaching is not homogeneously borne by the entire students population, there is room for systems of allocation of tasks and courses to teachers that match successful scholars with those students who benefit more from their knowledge and that minimize possible distortions. Second, I show that, while at the margin there is a trade-off between teaching and research, the overall effect is ambiguous: universities are also able to keep the best researchers under the new incentive regime and, since good researchers are also good teachers, teaching quality improves. Finally, I provide the first evidence on the correlation between research and teaching ability. This has important implications for the design of professors' incentives and hiring schemes. Policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that there is a large overlap between good researchers and good teachers.

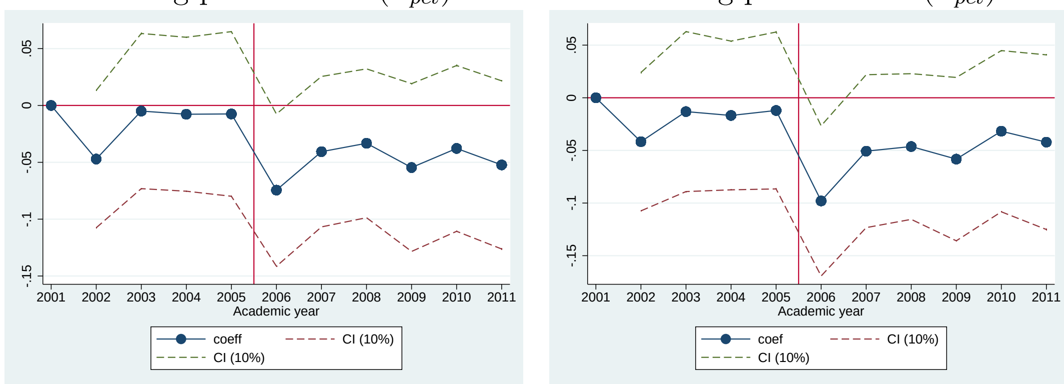
Figure 1: Research difference in difference graphs



Source: Web of Science and Google Scholar. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

Figure 2: Teaching difference in difference graphs

Internal professors teaching performance (α_{pct}) | Assistant professors teaching performance (α_{pct})

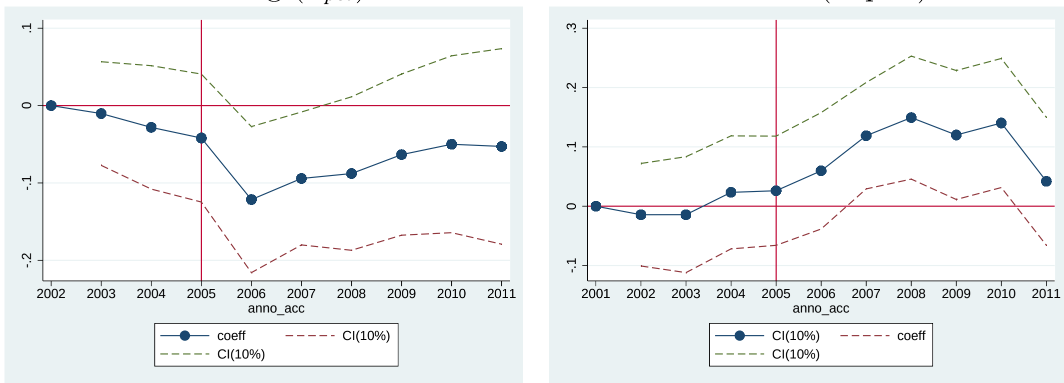


a b

Source: Bocconi student and teacher archives. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

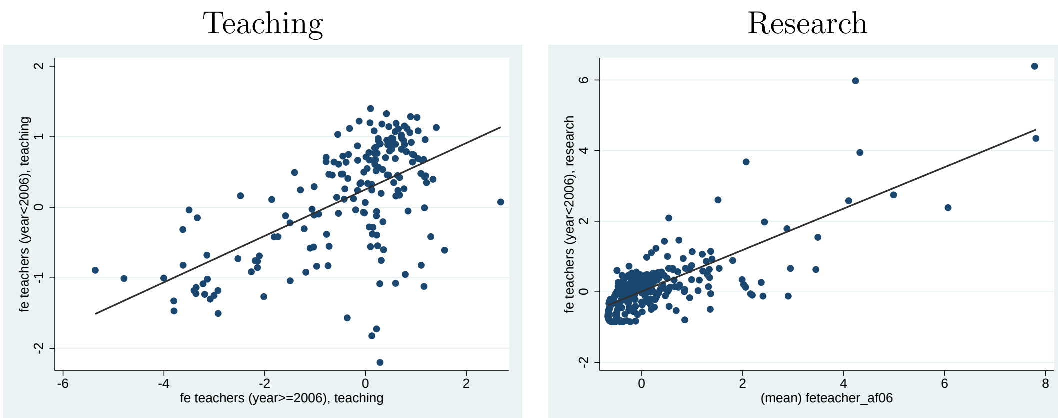
Figure 3: Alternative identification strategies graphs

control: full prof Teaching (α_{pct}) | control: Bologna prof Research (N pub)



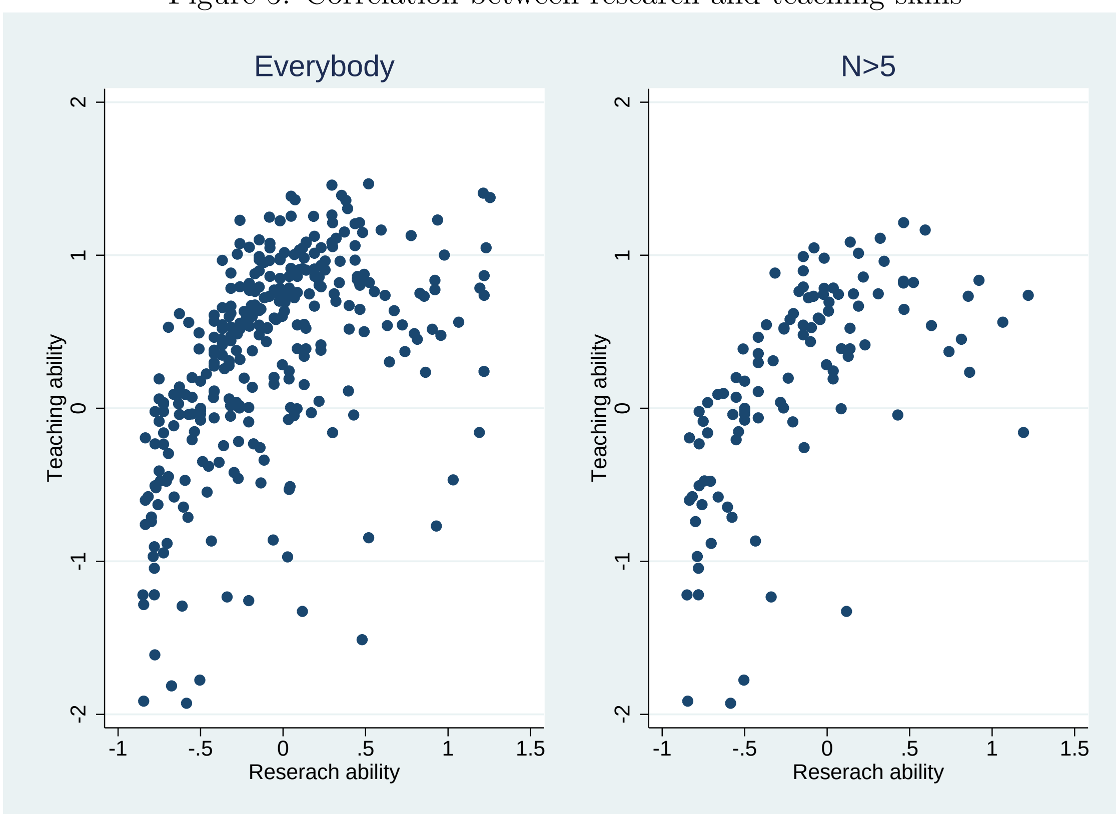
Source: Bocconi registers and Web of Science and Google Scholar. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

Figure 4: Robustness of teachers fixed effects



Source: Bocconi registers and Web of Science and Google Scholar. On the x axis=fixed effects based on the post 2006 sample; on the y axis= fixed effects based on the pre 2006 sample.

Figure 5: Correlation between research and teaching skills



Source: Bocconi registers and Web of Science and Google Scholar. Panel 1 plots $\widehat{\theta_p^t}$ and $\widehat{\theta_p^r}$ estimated on the entire sample of teachers. Panel 2 uses only the $\widehat{\theta_p^t}$ estimated on at least 5 observations.

Tables

Table 1: Descriptive statistics - Students

Variable	Mean [1]	Std. Dev. [2]	Min. [3]	Max. [4]
1=female	0.469	0.499	0	1
year birth	1985	3.249	1954	1993
1=italian	0.973	0.163	0	1
1=from Milan	0.246	0.431	0	1
HS grade	0.899	0.103	0.6	1
Exam grades	25.532	3.532	18	31
N		501189		

Source: Bocconi students' registers. The sample consists of students taking compulsory undergraduate courses between 2001-2011. High School (HS) grade normalized to be between 0 and 1 (pass if >= 0.6) for all counties. Exam grades refers to Bocconi exams.

Table 2: Types of teacher contracts

Description	category
Adjunct Professor	assistant
Researcher Bocconi	assistant
Assistant professor Bocconi	assistant
Assistant Professor (Job Market) Bocconi	assistant
Assistant Professor (Young Foreigners) Bocconi	assistant
1 year scholar Bocconi	assistant
2 year scholar Bocconi	assistant
3 year contract researcher Bocconi	assistant
Phd Student Bocconi	assistant
Assistant professor Bocconi senior	assistant
Researcher Bocconi	assistant
Full contract researcher Bocconi	assistant
Researcher Bocconi on leave	assistant
Associate professor Bocconi	associate
Full Professor Bocconi	full
Extraordinary professor Bocconi	full
Non academics (expert in the subject)	non academics
Associate professor other university	other univ
Associate professor Bocconi on leave	other univ
Temporary contract collaborator SDA[a]	other univ
Collaborator SDA	other univ
permanent contract collaborator Research centers	other univ
Full contract researcher SDA	other univ
Lecturer SDA	other univ
Lecturer SDA Senior	other univ
Full Professor other university	other univ
Full Professor Bocconi on leave	other univ
Associate professor other university	other univ
Full Professor other university	other univ
Researcher other university	other univ
Extraordinary professor other university	other univ
Visiting Professor Long Term	other univ
Visiting Professor Short Term	other univ

The big amount of contracts is due to the fact that identical contracts were having different names over the years.

[a] SDA is the Bocconi School of Managers. It offers MBAs and master course only. Faculty is hired and promoted according to different and independent standards.

Table 3: Descriptive statistics - Teachers

	Internal	External	Diff
Teachers' descriptives			
N teaching hours per class	38.91	33.91	5.47***
	(16.60)	(17.44)	(1.34)
Age	43.18	41.29	1.89**
	(9.45)	(7.80)	(0.77)
% female	32.27	34.25	-0.20
	(0.47)	(0.47)	(0.045)
Teachers' Department			
Accounting	14.8 %	20.8%	
Math/Stat	13.3%	24.6%	
Economics	20.2%	13.8%	
Finance	16.7%	7.4%	
Management	39.0%	33.5%	
Teachers' Contracts			
% Assistant prof	50.04%		
% Associate prof	10.45%		
% Full prof	12.65%		
% Non academic		9.61%	
% Other univ prof		9.11%	
% Lecturers	7.76%		

Source: Bocconi teachers' register. Standard deviation (columns 1 and 2) and standard errors (column 3) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 4: Random allocation

	Av. final hs grade[a]	Av. female	Av. from Mi	Sd final hs grade
1=int teacher	0.001	0.000	0.001	-0.002
	(0.001)	(0.000)	(0.003)	(0.002)
Teacher's Age	0.000	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
1=female teacher	0.001	0.000	0.000	-0.001
	(0.001)	(0.000)	(0.002)	(0.002)
1= course coordin	0.000	-0.001	0.003	-0.000
	(0.001)	(0.001)	(0.003)	(0.002)
N	3889	3889	3889	3889
course*year fe	Yes	Yes	Yes	Yes
F stat joint sign	0.75	0.95	0.39	1.58

The Table plots average classes characteristics (based on students' composition) on teachers' characteristics. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 5: Summary statistics-Research

	Overall	Post 2006	Pre 2006	Diff
		N publications		
Internal[a]	0.539	0.680	0.302	0.381***
sd	1.561			(0.061)
External	0.416	0.481	0.264	0.239***
sd	1.318			(0.042)
Diff	0.199	0.199**	0.037***	0.143**
	(0.199)	(0.100)	(0.071)	(0.074)
		N publications (Bocconi index[b])		
Internal[a]	0.814	0.927	0.625	0.336***
sd	2.328			(0.084)
External	0.575	0.634	0.437	0.251***
sd	1.944			(0.080)
Diff	0.293	0.293**	0.187*	0.085
		N working papers (Google Scholar)		
Internal[a]	1.506	1.692	1.193	0.526***
sd	2.583			(0.126)
External	1.052	1.159	0.809	0.343***
sd	2.278			(0.105)
Diff	0.533	0.533**	0.385***	0.182*
	(0.533)	(0.172)	(0.191)	(0.164)

Source: panel 1 and 2 Web of Science; panel 3: Google Scholar. Standard deviation (column 1) and standard errors (column 4, last row) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] This refers to the status in 2005.

[b] Publications are weighted in the same way Bocconi University assigns monetary incentives. I give weight=15 if articles are in journals considered by Bocconi as belonging to band "A+", weight=7 if journals are considered as belonging to band "A", weight=3 if belonging to band "B" and weight=1 if not belonging to any band. The index is computed as $\Sigma_i(weight_i * pub_i)/Nauthors_i$ where i is a publication published by professor p in year t.

Table 6: Effect on research performance

	[1]	[2]	[3]	[4]	[5]	[6]
	Dependent variable: N Pub				N pub$^{(1/2)}$	
Internal^a*post2006	0.206**		0.142**		0.081**	
	(0.100)		(0.070)		(0.034)	
Junior pr^a * post2006		0.224*		0.157*		0.087**
		(0.114)		(0.091)		(0.039)
Full pr^a* post2006		0.137		0.099		0.065
		(0.186)		(0.123)		(0.050)
N	5230	5230	5230	5230	5230	5230
	Dependent variable: N Pub (weighted by Bocconi)^b				N pub w$^{(1/2)}$	
Internal^a*post2006	0.315***		0.130		0.082**	
	(0.103)		(0.098)		(0.041)	
Junior pr^a * post2006		0.266**		0.154		0.094**
		(0.110)		(0.109)		(0.047)
Full pr^a* post2006		0.496**		0.064		0.050
		(0.231)		(0.174)		(0.060)
N	5209	5209	5209	5209	5209	5209
	Dependent variable: N working papers (Google Scholar)				N wp$^{(1/2)}$	
Internal^a*post2006	0.711***		0.148		0.091*	
	(0.166)		(0.139)		(0.052)	
Junior pr^a * post2006		0.492***		0.212*		0.120**
		(0.166)		(0.136)		(0.059)
Full pr^a* post2006		1.572***		-0.035		0.008
		(0.404)		(0.203)		(0.073)
N	5113	5113	5113	5113	5113	5113
Teacher fe	No	No	Yes	Yes	Yes	Yes

Additional controls: age, age squared, academic year fixed effects. Years between 2001 and 2011. Only professors included in the analysis on teaching. Junior professors are assistant and associate professors. The type of contract is defined according to the year 2005. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] This is the internal status in 2005

[b] Publications are weighted in the same way Bocconi University assigns monetary incentives. I give weight=15 if articles are in journals considered by Bocconi as belonging to band "A+", weight=7 if journals are considered as belonging to band "A", weight=3 if belonging to band "B" and weight=1 if not belonging to any band. The index is computed as $\Sigma_i(wight_i * pub_i)/Nauthors_i$ where i is a publication published by professor p in year t.

Table 7: Step 1: regression on students micro data.

Dependent variable: standardized exam grade	All
	[1]
HS grade	-3.704***
	(0.225)
HS grade^2	4.159***
	(0.131)
1=female	-0.051***
	(0.003)
1=italian	0.142***
	(0.013)
1=from Milan	0.074***
	(0.003)
N	501132

Additional controls: dummies for type of high school, dummies for the full interaction of classes and years (α_{pct}). Robust standard errors clustered by course-year in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 8: Descriptives for teaching performance

		α_{pct}			
		Overall	Post 2006	Pre 2006	Diff
Internal[a]	*mean*	-0.020	0.146	-0.197	0.343***
	sd	(0.632)			(0.024)
External	*mean*	0.074	0.239	-0.192	0.431***
	sd	(0.645)			(0.026)
Diff			-0.093***	-0.005	-0.088***
			(0.033)	(0.015)	(0.036)

α_{pct} estimated from Table 7, normalized to have mean 0. Standard deviation (column 1) or standard errors (column 4, last row) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Based on their status in 2005.

Table 9: Step 2: regression at teacher level - students' grades

	[1]	[2]	[3]	[4]
Internal[a]*post06	-0.011		-0.037**	
	(0.012)		(0.018)	
Junior pr[a]*post06		-0.014		-0.042**
		(0.013)		(0.020)
Full pr[a]*post06		-0.001		-0.023
		(0.016)		(0.022)
N	3889	3889	3889	3889
Teachers fe	No	No	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, class size, class average final high school grade. Junior professors are assistant and associate professors. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Based on status in 2005.

Table 10: Regression at the student level - heterogeneity by students' high school grades

	[1]	[2]	[3]	[4]
	Dependent variable: stud grade (std)			
	[1]	[2]	[3]	[4]
int[a]*post06	-0.037***		0.002	
	(0.014)		(0.016)	
jun[a] pr*post06		-0.045***		-0.005
		(0.016)		(0.017)
full[a] pr*post06		-0.009		0.028
		(0.020)		(0.022)
int[a]*post06*mid ability stud			-0.079***	
			(0.014)	
int[a]*post06*low ability stud			-0.097***	
			(0.020)	
jun[a]*post06*mid ability stud				-0.077***
				(0.015)
jun[a]*post06*low ability stud				-0.100***
				(0.021)
full[a]*post06*mid ability stud				-0.086***
				(0.022)
full[a]*post06*low ability stud				-0.086**
				(0.036)
N	346628	346628	346628	346628
Teachers fe	Yes	Yes	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Control set: teacher age, age squared, student gender, type of high school, whether Italian, whether from Milan. Ability based on final high school grade of students (normalized between 1 and 0, pass if >= 0.6): High ability (omitted)=between 1 and 0.9; middle ability = between 0.8 and 0.9; low ability0 below 0.8. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status as it was before 2006

[b] The number of observations is lower because Bocconi collected students evaluations in only a subsample of courses for the years 2004/2005 and 2005/2006.

Table 11: Robustness checks for the teaching regression

	no 09-10-11 [1]	also lecturers [2]	include switches [3]	weight by h. taught [4]
Internal[a]*post06	-0.037*	-0.034*		-0.035*
	(0.020)	(0.018)		(0.021)
Lecturer[a]*post06		-0.047		
		(0.042)		
Internal[b]*post06			-0.027*	
			(0.016)	
N	2848	4201	3889	3889
Teachers fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, teacher experience in Bocconi class size. Column (1) excludes the years when teaching incentives were also in place; column (2) includes lecturers and specifies a different treatment effect for lecturers; column (3) includes switchers and teachers fixed effects; column (4) weights professors by number of teaching hours. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status as it was before 2006

[b] Contemporaneous status

Table 12: Robustness checks for the teaching regression 2

Dep var:	Grading α_{ptc}		1=course coordin[a]	1=Num of taught h[b]
	[1]	[2]	[3]	[4]
int*post 06	-0.045**	-0.042**	0.025	0.671
	(0.020)	(0.020)	(0.037)	(1.084)
int*post 06*obj[c]	0.024			
	(0.047)			
int*post 06*math dep[d]		0.017		
		(0.046)		
N	3889	3889	3889	2989[e]
Teachers fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, dummies for teacher experience in Bocconi. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] 1=whether professor p in year t was the course coordinator

[b] Tot n of teaching hours in year t by professor p

[c] Objective if the name of the course includes the words *"math"*, *"stat"*, *"quantit"*

[d] Math if the teacher belongs to the math and statistics departments

[e] N of observations at the teacher-year level (if a teacher teaches more than one courses n of teaching hours are summed)

Table 13: Robustness checks for the teaching regression 3: average class grades

	sd av. class gr internal	sd av. class gr external
2001	0.390	0.434
2002	0.283	0.379
2003	0.390	0.453
2004	0.415	0.375
2005	0.423	0.431
2006	0.407	0.477
2007	0.375	0.400
2008	0.450	0.349
2009	0.406	0.442
2010	0.428	0.425
2011	0.468	0.404

[a] This is the standard deviation of average class grades within courses (of classes that sit the same exam).

Table 14: Heterogeneity by teachers' research skills

	Teaching	Research		
Dep. var	α_{pct}	n pub	n pub weight	n wp (google)
	[1]	[2]	[3]	[4]
int*post 06* ability q 1	-0.087**	0.146***	0.185**	0.296***
	(0.037)	(0.042)	(0.073)	(0.113)
int*post 06* ability q 2	-0.036	0.200***	0.351***	0.166
	(0.041)	(0.070)	(0.102)	(0.172)
int*post 06* ability q 3	-0.035	0.184	0.197	0.095
	(0.042)	(0.140)	(0.257)	(0.254)
N	3770	6281	6264	6082

Additional controls: age, age squared, all double interactions, teacher fixed effects, year fixed effects. Ability based on tertiles of the teachers fixed effects on research obtained from estimating euqation 8. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 15: Teaching and research skills

	Dep. var= Teaching Fe (θ_p^t)			
	everybody	N>5[a]	everybody	N>5[a]
	[1]	[2]	[3]	[4]
Research FE (θ_p^r)	0.715***	0.795***	0.542***	0.640***
	(0.067)	(0.102)	(0.062)	(0.094)
N	313	109	313	109
Controls	No	No	Yes	Yes

Additional controls: age at entry (linear and squared), gender.
[a] N>5 is referred to the n of observations over which is estimated the teacher fixed effect in the teaching quality regression (for the research quality regression N=10 for every teacher). Robust standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 16: Sorting out

	Fixed Effects all		pre 06 Fixed Effects	
Dep Variable:	Research Fe	Fe Teaching	Fe Research	Fe Teaching
	(θ_p^r)	(θ_p^t)	(θ_p^r)	(θ_p^t)
	[1]	[2]	[3]	[4]
1=exit after 2006	-0.133**	-0.113**	-0.099*	-0.100
	(0.063)	(0.054)	(0.055)	(0.074)
1=exit pre 2006	-0.044	0.023	-0.008	-0.040
	(0.054)	(0.034)	(0.046)	(0.036)
N	345	352	232	232

Columns 1 and 2 use fixed effects estimated over the entire time period. Columns 3 and 4 use fixed effects estimated only before the change in incentives. Excluding those exiting because retiring, omitted category=those staying. Additional controls: dummies for year of entry, gender, age at entry, age at entry squared. Robust standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 17: Sorting in

Dep. Variable:	Research Fe (θ^{r_p}) [1]	Fe Teaching (θ^{t_p}) [2]
1=entry after 2006	-0.051	0.049
	(0.091)	(0.056)
tr y entry	0.009	0.099***
	(0.011)	(0.013)
tr y entry sq	0.001*	-0.000
	(0.000)	(0.000)
N	350	352

Excluding those exiting because retiring, omitted category=those staying. additional controls: dummies for year of entry, gender, age at entry, age at entry squared. Columns 3 and 4: omitted category= entry before 2006, additional controls=time trend of year of entry (linear and squared), age at entry (linear and squared and triple), gender. Only for teachers entered after 2000. Robust standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 18: Summary statistics on publications Bocconi and Bologna professors

	Bologna	Bocconi	Difference
	Junior prof		
N publications	0.201	0.417	-0.217***
	(0.018)	(0.022)	(0.033)
	Senior prof		
N publications	0.280	0.481	-0.201***
	(0.030)	(0.043)	(0.051)

Based on faculty composition in 2005, refers to yearly publications between 2001 and 2011. Source: Web of Science. The junior professors category includes assistant and associate professors. Standard deviation (column 1 and 2) or standard errors (column 3) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 19: Alternative identification strategies - Research

Dep. var	Number of publications				
Contr gr	External professors age groups		Bologna professors		
	< m age (43)	> m age (43)	All	Jun	Full
	[1]	[2]	[3]	[4]	[5]
Internal[a]*post 06	0.170*	0.119			
	(0.096)	(0.124)			
Bocconi[a]*post 06			0.162**		
			(0.064)		
Jun bocc[a]*post 06				0.221***	
				(0.080)	
Ord bocc[a]*post 06					0.051
					(0.107)
N	3119	2111	4497	3063	1434

Additional controls: age and age squared of teachers, year fixed effects. column (1) and (2) use as control group external teachers in the same age group (< or > median age of 43), columns (3), (4) and (5) use as control group professors from Bologna University. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status based on the year 2005.

Table 20: Alternative identification strategy - Teaching

	Dep Var: α_{pct}			
Contr gr:	External prof age groups		Prof just became tenured	
	< m age (43)	> m age (43)		
	[1]	[2]	[3]	[4]
Internal[a]*post06	-0.061*	-0.034		
	(0.032)	(0.029)		
No full[a]*post06			-0.221***	-0.042*
			(0.052)	(0.025)
N	1958	1931	2068	2068
Teachers fe	Yes	Yes	No	Yes
Year*course*deg fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, dummies for year of arrival in Bocconi. Only internal teachers (in 2005). Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status based on the year 2005.

2.8 Appendix

Given the exponential utility function and normality of ϵ_i, the agent receives certainty equivalent

$$CE = b_r\alpha_r e_r + b_t\alpha_t e_t + s - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t - \frac{\eta}{2}(b_t^2\sigma_t + b_r^2\sigma_r) \tag{11}$$

The first order conditions obtained from maximizing the expected utility of the agent with respect to e_r and e_t are:

$$\alpha_r b_r = e_r + \delta e_t; \quad \alpha_t b_t = e_t + \delta e_r \tag{12}$$

and the optimal (internal) solutions are:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{13}$$

Therefore, taking the partial derivatives with respect to b_r, I get:

$$\frac{\partial e_r^*}{\partial b_r} = \frac{\alpha_r}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r} = -\frac{\delta\alpha_r}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{14}$$

To show the results stated in Proposition 2, I take the derivatives also with respect to ability:

$$\frac{\partial^2 e_r^*}{\partial b_r \partial\alpha_r} = \frac{1}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r \partial\alpha_r} = -\frac{\delta}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{15}$$

Chapter 3

Parents, Schools and Human Capital Differences across Countries

3.1 Introduction

According to international standardized tests, there are large and persistent cross country differences in the performances of students of similar age. East Asian countries like Korea, Japan, China and Singapore consistently position themselves at the top of international rankings, while the relatively disappointing performance of several Southern European and Latin American countries has become a hot topic in the public debate. This is true, though to a slightly different extent, across all subjects commonly tested in these international studies and across all years for which these comparisons are available.

While these facts have been well known at least since the release of the first results of the PISA test, they have recently received renewed attention in light of an emerging literature that puts them at the centre of the discussion on cross country differences in economic performance. Kimko and Hanushek (2000), Hanushek and Woessmann (2012a) and Schoellman (2012), among others, argue that average years of schooling, the standard proxy in the growth literature for the quantity of human capital, does a rather poor job of measuring differences across countries in terms of the knowledge embodied in their workers, while standardized tests allow to capture differences in terms of human capital quality which turn out to have much greater explanatory power for differences in GDP.

Given the increasing role these cross-country gaps in standardized tests play in the growth literature, it becomes important to understand where they come from. Most of the public and academic debate on this issue tend to use (and argue in favour of) an interpretation of PISA scores as measures of school quality. The popular press is rich of anecdotes on the severity and depth of school curricula in some Asian countries, which introduce students to challenges often quite demanding for their age. On the other hand, there is some evidence (once again, mostly anecdotal) of important differences in terms of "parental inputs", a broad category which includes parenting styles and parents' attitudes and beliefs towards education. The international best-seller by Chua, Amy (2011) coined the expression "Tiger Mother" to describe the rather strict way in which some Asian parents raise up children, pushing on academic excellence and very long studying hours.

In this paper we aim at shedding some light on the importance of parental inputs in PISA test scores. We show that a substantial share of cross-country gaps are reproduced across second generation immigrants educated in the same school as children from other countries. Discriminating between parental inputs and the "school" component has rather important implications, since it determines whether Western countries aiming at improving their students' performance should consider imitating some of the characteristics of the East Asian school system, or whether the explanation for achievement gaps lies in deeper cultural factors, perhaps harder to affect for policy makers. With this in mind, our objective is to investigate how much of the cross country variation in test scores can be attributed to differences in parental inputs, and what is the nature of these differences in parental inputs in the first place.

It is important to clarify that our focus in this paper is on the cross-country variation in average test scores only, as opposed to the variation across students within a given country. We are therefore after a country-specific parental component, which captures a set of practices, inputs, attitudes and beliefs that on average belong to parents of a given nationality. Of course parents differ on a wide array of dimensions even within countries, and it is likely that some of these differences will be important determinants of individual students' performances in standardized tests. However, for the study of cross country gaps in average scores, and cross country differences in GDP growth, all that matters are gaps in terms of average parental inputs across countries.

Such analysis is obviously complicated by the fact that parental inputs are typically difficult to measure, and, even when proxies are available, it is unclear how to separately identify their effect from the quality of the school system. In this paper we wish to overcome these difficulties by studying test scores for second generation immigrants. We identify the importance of country-specific parental inputs by comparing the performance of students born and educated in a given country and in the same school, but with parents of different nationalities. Since factors such as the school curriculum, teaching style and school infrastructure (as well as other individual level characteristics) are being kept fixed in this comparison, we can attribute any residual difference to inputs received by their parents. We then show that the results from this simple empirical exercise can be used to decompose the cross-country variation in test scores between different sources, shedding light on the nature of these gaps. Our paper, therefore, shares the spirit and the approach of several studies that look at first and second generation immigrants to identify the importance of "portable" cultural and familiar componentsfor various different outcomes (the so-called "epidemiological approach" (Alesina et al., 2013; Fernandez, 2011; Alesina and Giuliano, 2010; Fernandez and Fogli, 2009; Giuliano, 2007)). As we discuss at length, our approach is unlikely to be biased by a pattern of differential selection of emigrating parents from different countries. Indeed we check that our relation of interest is not spuriously generated by the fact that emigrants from high PISA score countries are more positively selected into migration than emigrants from low PISA score countries. If anything, we find a pattern of selection which seems to go against our findings.

Our results point towards a substantial role of the parental component. We find the performance in the PISA test of second generation immigrant students, living in the same country and studying in the same school, is very closely related to the one of natives[1] from the country of origin of their parents. In particular, second generation immigrants from high PISA countries score better than their peers from low PISA countries, even when they are observed in the same school and even if their parents have the same level of education and type of occupation. This pattern is present also when we focus on a different schooling outcome from a different sample, such as the probability of grade repetition in the US Census; once again, the best performing second generation immigrants are those whose parents come from countries where natives are particularly successful in standardized tests. Taking our estimates at face value, we find that at least 12% of the total cross-country variation in test scores can be accounted for by differences in parental inputs. Their contribution is substantially higher when we look at the gaps between specific countries: for example, at least 33% of the out-performance of East Asian and Souther European countries is persistent across second generation immigrants, suggesting that parental practices play a predominant part

[1] Throughout the paper, we call natives those students born in the country where they are taking the test and whose parents are born in the same country as well. Students born in a country different from the one where they are taking the test are excluded from all the analyses that follow.

in explaining this gap.

We then move to explore more in detail the nature of these differences in terms of parental inputs, making use of both the PISA and US Census data as well as of several other sources. We first show that the relationship between the performance of a second generation immigrant and the average score in the parents' country of origin is weaker for parents that are more educated. This suggests that what drives our results is not related to the quantity or quality of education received by parents in their home country. Moreover the relationship weakens if parents have spent more years in the host country, suggesting the importance for school performance of country-specific "cultural" traits, that are progressively lost by immigrants as they integrate in their new host country. Then we look at detailed time use surveys on immigrants in the US, to investigate whether differences in parents and students' observable practices can help to explain gaps in performance between children of different nationalities. Our results suggest that this is a promising avenue for further exploration, since parents from high PISA countries systematically spend more time on various forms of childcare, while their children spend more time studying and in related activities.

Beside contributing directly to the previously mentioned growth literature on cross country differences in human capital quality, our results speak to the literature that studies the role for the performance of immigrant children of various attributes of the destination countries or of the home countries (for instance Dronkers and de Heus (2012),Dustmann et al. (2012) and Jerrim (2015)). We contribute to this literature first by extending the analysis to about 40 host countries and 60 countries of origin instead of focussing on specific communities (Dustmann et al. (2012) study Turkish immigrants and Jerrim (2015) East Asian immigrants); second by quantitatively decomposing how much of the cross country differences in test scores is actually attributable to the country of origin of the children's parents and how much to the destination country; third by providing some suggestive evidence of the mechanism behind the relationship. Moreover, our findings are connected to the literature examining schooling and labor market outcomes as a function of ethnicity. In his seminal work, Borjas (1992) uses data from the General Social Survey (GSS) and the National Longitudinal Survey of Youth (NLSY) to show that the quantity (not the quality) of education in the ethnic environment of parents, what he calls "ethnic capital", plays a role in the human capital accumulation process of the following generations in the US. Dustmann et al. (2010) study the evolution of the attainment gap between white British born and ethnic minority pupils throughout compulsory schooling. Finally, the finding that parental inputs are important determinants of school performance speaks to the literature highlighting the role, for children educational attainment, of transmittable cultural values such as attitudes towards school, aspirations and non cognitive skills (Heckman and Rubinstein, 2001; Brunello and Schlotter, 2011; Behncke, 2012; Borghans et al., 2008; Carneiro et al., 2007).

The paper is structured as follows. Section 3.2 describes the data. Section 3.3 shows empirical evidence on the performance of second generation immigrants as a function of their parents' country of origin. Section 3.4 addresses the possibility that our findings are driven by different forms of selection. Section 3.5 makes use of these results to quantify the overall importance of the parental component for cross-country differences in test scores, while Section 3.6 explores more in detail the possible mechanisms behind our results. Finally, Section 3.7 concludes.

3.2 Data

Our main data come from the 2003, 2006, 2009 and 2012 waves of the PISA test. PISA is a triennial survey of the knowledge and skills of 15-year-old children, explicitly designed to allow comparisons across countries. Since 2003, 73 countries have administered at least one wave of the test, covering all OECD members as well as some partner countries. Typically, each country selects between 4,500 and 10,000 students through a two-stage stratified sampling technique, where a random sample of at least 150 schools enrolling 15-year-old students is drawn first, and then 35 students within each school are randomly selected to take part to the test. Throughout the analysis, we make use of the sample weights provided by the OECD.

The PISA tests cover three different subjects: reading comprehension, science and mathematics. Neither students nor teachers know the result of the test at the end, so these are rather low stake tests for students. Since each student is tested on a random subset of questions, test results are not presented as point estimates, because they will not be easily comparable, but rather as "plausible values". The OECD estimates for each student a probability distribution of test scores based on their answers, and randomly draws from it five values (defined "plausible values"), see OECD (2011) for details.[2] PISA scores are internationally standardized to have mean 500 and standard deviation 100 across OECD countries. We further standardize them to have mean 0 and standard deviation 1 for each subject, across all available waves.

[2]Throughout the analysis, when using the PISA data, we compute the standard errors of our estimates by using the unbiased shortcut procedure described in OECD (2009) and followed by the literature (i.e. Dustmann et al. (2012)), to take into account that we use plausible values for the test scores, as the OECD only provides plausible values.

As it is well known, results for all subjects vary greatly across countries. In an attempt to summarize this variation, Table 1 shows the average score within a number of broadly defined geographical regions. The superior performance of East Asian countries is particularly strong in mathematics, but the ranking across regions is quite stable across different subjects. The magnitude of the gaps is quite striking. According to OECD (2012a), a gap of 0.4 in this scale corresponds roughly to what is learned in an average year of schooling. It follows that East Asian students have a more advanced knowledge of mathematics which corresponds to two additional years of schooling compared to Southern Europe, and to almost four additional years of schooling compared to Latin America and Middle East/North Africa.

The PISA data include a Student Questionnaire in every wave, which provides basic demographic information on students and parents. The exact country of origin of parents is, however, not available in all participating countries questionnaires and for the wave 2000.[3] On top of this, for some countries and waves further information is available from the School Questionnaire and the Parent Questionnaire. In particular, we use the School Questionnaire to construct some measures of school quality, and the Parent Questionnaire to get additional information on parents' age and education.

The final sample includes 43486 second generation immigrants on the mother side and 43728 on the father side, from 49 and 48 different countries of origin and distributed across 39 host countries. Descriptive statistics for second generation immigrants on the mother side are provided in Table 2. Table 3 displays the number of available observations for each country of origin that can be included in our decomposition exercise and the main receiving country for each country of origin. Sample sizes vary greatly, and for some countries of origin we have only a few second generation immigrants. To account for this, we weight countries of origin by the number of second generation immigrants in the sample when constructing averages for broad geographical regions.[4] Table 4 shows the same set of information but referred to host countries. Our final sample includes only host countries for which we have information on the parents' country of origin. Each destination country receives immigrants from, on average, 6 different countries of origin.

The second source of data is the Integrated Public Use Micro data Series (IPUMS) created

[3]Individual countries have some flexibility on how to classify parents' country of origin. While most have indicators for each country, some group small countries in broader categories. We construct a set of countries/regions consistently defined over time, and drop observations for second generation immigrants whose parents do not come from any of these countries. See the data Appendix for the details.

[4]Moreover, in all graphs plotting estimates relative to individual countries of origin we weight observations by the inverse of the corresponding standard error.

by the US Census Bureau. The IPUMS consists of individual and household level data from the decennial census in the US and includes nearly all the details originally recorded by the census enumerations. We use the 1% samples from the 1970 and 5% sample from the 1980 censuses. Even if IPUMS has little information on children's outcomes, it does, however, contain information on each individual's exact grade attending at school.[5] We follow Oreopoulos and Page (2006) in combining this information with children's age to construct an indicator of whether or not each student has repeated any grade. As pointed out by Oreopoulos and Page (2006), grade repetition is a widespread phenomenon in the United States and is correlated with many commonly used measures of educational achievement and socioeconomic success. We classify a child as a repeater if her educational attainment is below the mode for her state, age, quarter of birth, and census year cell. Following Oreopoulos and Page (2006), we focus on children between the ages of 8 and 15, since children younger than 8 have not had many opportunities to repeat a grade, and children older than age 15 might have left home already or dropped out of school. To adjust for the fact that older sample members have had more opportunities to repeat a grade, and to adjust for possible gender differences in grade repetition, all regressions include controls for age dummies and gender. Moreover, we experimented with several alternative definitions of grade repetition and our results are robust throughout.[6]

The final sample includes 53361 second generation immigrants on the mother side and 46685 on the father side, from 61 different countries of origin. Descriptive statistics for second generation immigrants on the mother side are provided in Table 6.

For Section 3.6.2, we use the ATUS-US Time Use Survey to analyze how immigrant parents and their children spend their time. We pool together all waves between 2002 and 2013. The ATUS survey was administered only to one person per household, chosen randomly among all individuals at least 15 years old. We use both data on parents and children, where children are those individuals between 15 and 18 years old. For parents, we construct a variable measuring the total time (in minutes) spent on child care on the previous day, and three subcategories that split total child care in educational, recreational and basic activities.[7] Descriptive statistics on parents' and students' time are displayed in Tables 15 and 17.

Finally, we rely on several other sources to construct our controls at the level of parents' country of origin. We take GDP per capita in 2006 from the PWT, average years of schooling

[5]This information is only available until 1980, which prevents us from using more recent years.

[6]Results are available upon request.

[7]We follow Aguiar and Hurst (2007) for the construction of these variables. See Appendix 3.8 for the details.

for different demographic groups from Barro and Lee (2013), measures of school quality from Bartik's "School Quality" dataset, various answers from the World Value Survey to proxy for cultural differences and data on the religion composition in 1970 from Barro and McCleary.

3.3 Reduced Form Evidence

In this section we examine whether the school performance of second generation immigrants is related to the one of natives[8] in the parents' country of origin. Throughout the section, we focus our analysis on second generation immigrants on the mother side only. This is done only to simplify the exposition, and alternative specifications, available upon request, show that our results hold without exception when we look at second generation immigrants on the father side or at the whole sample of second generation immigrants and natives.[9] We present results for the PISA and the US Census samples in turn.

3.3.1 PISA

Let T^m_{icst} denote the PISA math[10] score in year t of child i, studying (and born) in country c and in school s (or country c), whose mother was born in country m.[11] We start from the following specification:

$$T^m_{ist} = \theta_0 + \theta_1 T^m + \theta_2 X^m_{ist} + \theta_3 Z^m + \theta_{st} + \varepsilon^m_{ict} \tag{1}$$

[8]Throughout the paper we use the term natives to refer to the group of children born and educated in a certain country, whose mother/parents were born in the same country.

[9]This more complete specification will be used for our decomposition in Section 3.5

[10]The results are similar for the Reading and Science tests (results available upon request). The Math test is often preferred for international comparisons for the relative easiness of defining and quantifying a common set of expected skills (Hanushek and Woessmann, 2012a).

[11]Throughout the paper, subscripts refer to the location and characteristics of students, while superscripts refer to the country of origin of parents.

where T^m is the average score of native students in the mother's country of origin[12][13], X_{ist}^m is a vector of individual characteristics of student and parents, Z^m are controls at the mother country of birth level, θ_{st} is a country or school (depending on the specification) fixed effect, specific for each PISA wave and ε_{ict}^m is an error term. The main coefficient of interest is θ_1, which captures the relationship between a given second generation immigrant's performance and the average score of native students in country m.

Here, T^m is used as a proxy for the bundle of characteristics of parents born in country m which affect the school performance of their children. The average test score in a given country reflects a combination of school quality, economic, cultural and institutional factors. However, by analyzing children educated in the same country/school (through the use of θ_{st}), who differ just because their parents come from different countries, we disentangle the part of their tests scores related to institutions/school system from the part related to parental inputs. The main worry is of course that omitted inputs for students' performance might be correlated with T^m, i.e. that, for example, second generation immigrants whose parents come from high PISA countries might receive higher investments in their educational development for reasons unrelated to their parents' nationality. The school fixed effect takes care of the possibility that they might attend schools of higher quality, given that we are comparing students within the same school. On top of this, we also control for parental characteristics which might be correlated with human capital investments on children and with PISA scores, such as parental education and socioeconomic status.[14] In this way we are able to understand how much of the effect comes from differences in parents' observable characteristics (education level, type of schools) and how much from differences in an unobservable, country-specific, component. The possibility that our results are driven by differential selection on unobservables of second generation immigrants will be discussed in great detail in Section 3.4.

Table 5 shows our main results. The sample is limited to second generation immigrants on the mother side, and a dummy is included in all specifications to control for whether the father

[12]The average score is computed across all available waves, applying the provided sample weights

[13]In this setting the dominant information is found in cross-country variation. A panel data approach would be of difficult implementation because the main regressor is indeed persistent over time (see Kimko and Hanushek (2000)), and short run shocks and variations in PISA scores are likely to be caused by cohort effects rather than significant changes in the cultural and educational environment. Moreover, PISA tests are available for few points in time and for recent years only. This prevents us from including country of origin fixed effects in this specification.

[14]Results with socioeconomic status as controls are not show here, and are available upon request. Information on parental age is available only for country and waves for which the Parent Questionnaire was administered. Our results are robust to the inclusion of these controls in this subsample.

is also foreign born.[15] We proceed by progressively adding controls. Column (1) of Table 5 displays the raw correlation between PISA scores of second generation immigrant students whose mother comes from country m and the average PISA score of natives in country m. It is strong and highly significant: if we compare two second generation immigrant students, with the mother of the first coming from a country where students score a standard deviation higher than the mother of the second, we see the former doing better than the latter by 66% of a standard deviation. The coefficient shrinks when we restrict the comparison to students that are observed in the same country (Column 2) and, especially, in the same school (Column 3), but is still positive and significant. The difference in the size of the coefficient between the first two specifications and Column 3 is quite illuminating, since it suggests that mothers from high PISA countries tend to send their children to better schools. We will show further evidence of this and discuss some implications for our empirical exercise in Section 3.4.

The specification in Column 4 adds controls for parental education, with the coefficient of interest being hardly affected. This finding is useful for the interpretation of the mechanisms behind our results: it suggests that the estimate of θ_1 is not driven by some parents' unobservable skills (like IQ for instance), otherwise we would expect these unobservables to be correlated to parental education, and therefore we expect the inclusion of this last variable to matter a lot for our coefficient of interest. What drives our result seems to be something not correlated with parents' education level. A possible concern, though, is that θ_1 is not much affected because parental education for immigrant parents is measured with error. Indeed, the estimated coefficients for these variables are quite small and, with the exception of the tertiary education dummy for fathers, not statistically significant. Moreover, the presence of some measurement error is quite realistic, given that PISA questionnaire are filled in by students, who may have difficulties reporting their parents' educational level, especially if parents were educated in a different country. We exploit the fact that for countries and waves where the Parents Questionnaire was administered, parents were asked to report their education as well. We therefore instrument the mother's and father's educational levels, as reported by children, with those reported by the parents themselves. Since the sample that allows this exercise is considerably smaller, we focus on the specification that includes both natives and second generation immigrants on either parent's side.[16] The results (available upon request) show that, while there is some degree of measurement error, the coefficient of

[15]This specification therefore ignores the variation in parental inputs associated with the country of origin of the father. As mentioned earlier, specifications that focus on fathers or that include the whole sample of second generation immigrants and natives give very similar results.

[16]As discussed later in the paper, in all specifications on the whole sample we include an interaction term between a given parent's native status and the average score in his or her country of origin, such that the coefficients of T^m and T^f are identified only out of second generation immigrants.

interest θ_1 does not vary much in magnitude between the OLS and the IV specifications.

Finally, the last two columns of Table 5 show that results are not driven by the particularly good performances of Asian students, since the coefficient is robust to the inclusion of continent fixed effects and to the exclusion of Asian mothers.

Figure 1 summarizes the results presented in this section. The left panel plots the average score of second generation immigrants, studying abroad but whose mother is from a given country against the average score of natives in that country, displaying graphically the strong and positive relationship described in Column 1 of Table 5. The right panel shows the correlation weakens but is still positive and significant when we clean the scores of second generation immigrants from the effect of differences in observable characteristics.[17]

3.3.2 US Census

In this section we apply a similar specification as in equation (1) on the US Census data, using a dummy equal to one if a child has never repeated any grade as our dependent variable. We notice that this outcome, while still related to school performance, captures quite a different dimension compared to the PISA score, given that the variation in this case comes only from the bottom part of the distribution (more than 80% of the students in the sample has never repeated a grade, as shown in Table 6). On the other hand, while the PISA dataset contains only 15 year old children, the US data allows us to look at students between 8 and 15 years old. We therefore find quite noteworthy that our results generalize to this sample as well.

The US Census does not contain any information on the particular school children are attending, making it impossible to compare second generation immigrants in the same school, as we did for Table 5. In an attempt to capture some of the differences across educational systems within the US, we control for State and Commuting Zone[18] fixed effect. However, the US Census provides us with precious information on parents' immigration history, so that we can control for the number of years passed since the mother has first migrated to the US. This is important to inspect the mechanisms behind our results. On top of this we

[17] This graph is plotting the estimated θ^m's from the regression

$$T^m_{ist} = \theta_0 + \theta_m + \theta_2 X^m_{ist} + \theta_{st} + \varepsilon^m_{ist}$$

where θ_m are mother's country of birth fixed effects.

[18] Commuting Zones are constructed following Autor and Dorn (2013).

can also control for parents' age and family size.

Table 7 shows our main results. Once again the coefficient on T_m^m is positive and significant, and does not change much in magnitude when controls for parental education and years since migration are included. According to column 3, the most complete specification, an increase of a standard deviation in the PISA score of students in the mother's country of origin is associated with a higher probability of not having repeated any grade by 3 percentage points (4% over the average). This effect is not trivial, given that, as mentioned earlier, most students do not repeat any grade. As for the PISA specification, the result is robust to the inclusion of continent fixed effects and the exclusion of Asian parents.

3.4 Selection

As our analysis relies on emigrant parents to make inference on all parents of a given nationality, an obvious concern is represented by the fact emigrants are not a random sample of the population, and might be selected on unobservable characteristics that also matter for children's school performance.

What type of selection should we worry about? Figure 2 displays various possibilities. In these plots the solid line represents the actual relationship between the score of second generation immigrants and the one of natives from the parents' country of origin, i.e. the relationship that we would be able to observe in a world where emigrant parents were randomly selected from the population. The dashed line represents instead what we would observe in our data, assuming different pattern of selection into emigration. The first panel depicts the case where the type of selection into emigration (as measured by the gap between the two lines) is the same across countries of origin with different PISA scores: if this is the case, only the estimate of the intercept of our regression will be biased. In the second panel we have the case where parents emigrated from countries with high PISA scores are more positively selected than parents emigrated from countries with low PISA scores, while in the third panel we have the opposite case. These patterns of differential selection would lead to a biased estimate of our coefficient of interest, and in particular the case depicted in the second panel could rationalize the findings of the previous sections.

While the main threat to our approach is represented by differential selection on unobservables, it is useful to verify whether emigrant parents are differentially selected on observable characteristics. The idea here is that unobservables that positively affect children's school performance are likely to be positively correlated with some observable characteristics, like parents' education and socioeconomic status. We can therefore somehow alleviate the concerns on differential selection by showing that the relative "quality" of emigrants compared to stayers is not higher for high PISA countries.

For each emigrant parent we construct a measure of selection given by the ratio between his or her years of schooling and the average years of schooling of non emigrant parents' from the same country.[19] Figure 5 plots the average of this measure of selection across mothers' countries of origin against the average score of native students in those countries. If anything, the relationship seems to be negative, suggesting that emigrants from high PISA countries might be more adversely selected than emigrants from low PISA countries (panel 3 of Figure 2).

Moreover, Table 8 shows results of a regression of our measure of selection of emigrant parents on the average PISA score in their country of origin, controlling for country (columns 1 and 3) and school (columns 2 and 4) fixed effects. The pattern is rather similar for mothers and fathers: we see rather weak evidence of negative selection within host countries, and stronger evidence within schools.

We interpret this pattern of negative differential selection within schools as consistent with the view that there is assortative matching between the quality of parental inputs and the quality of the school attended by children. The logic is as follows: when we observe two students in the same school, with one of them having received higher quality parental inputs (which in our setting means having parents born in high PISA country), then it must be that the other student (or parents) are "better" on some other (unobservable) characteristics, otherwise the two students would not be in the same school to start with.

Another issue, which is not related to selection into immigration but to selection into schools/host countries, is connected to the inclusion of θ_s fixed effects in our regressions and with the interpretation of the coefficient θ_1. If, for instance, parents from high PISA countries select better schools/host countries than parents from low PISA countries and we think that the ability/willingness to select good schools should be considered as part of the parent component, then the inclusion of a school fixed effects will absorb it. In what follows we try to quantify how much does the inclusion of θ_s change the estimates of θ_1 by exploring

[19] We construct a mapping between the ISCED classification of educational levels and equivalent years of schooling by using the country specific conversion table in OECD (2012b).

how school selection is related to the type of parents' country of origin.

In order to provide evidence for whether, everything else equal, second generation immigrants with parents from high PISA countries attend better school, we use the proxies for school quality we constructed from the information available in the School Questionnaire. Table 9 shows that, after controlling for country fixed effects and the usual observable characteristics, a higher PISA score in the country of origin of the mother is associated with schools where natives score better in the PISA test, no matter whether we take the raw average (column 1) or we clean it from observable characteristics (column 2), where admissions are more likely to be based on academic records, the proportion of teachers with at least some tertiary education is higher and the proportion of students dropping out is lower.[20]

To summarize, second generation immigrants from high PISA countries seem to be attending better school, and this is likely to generate a pattern of differential selection that biases our result downwards when we use within school variation. What if we limit ourselves to within country comparisons? In that case, the effect of these differences in the quality of the school attended would be erroneously attributed to the country of origin of the parents, conflating our coefficient of interest. For this reason, we view the specification with school fixed effects as a lower bound for the importance of parental inputs, we will show also results from the specification with host country fixed effects only which we can interpret, loosely speaking, as an upper bound.[21]

A final concern about selection is that immigrant parents from high PISA countries may be better at selecting host countries (or schools) where they children may better integrate and perform at school. We check whether immigrant from high PISA countries are systematically located in countries which are culturally and linguistically closer to their country of origin. To explore this possibility, Table 10 explicitly looks at the linguistic dimension. In column 2 we add to the baseline regression of column 1 a dummy variable that takes value of 1 for all students that declare to speak a foreign language at home (This observation is available only for part of the sample). While the coefficient on this newly added control is, as expected, negative and significant, our main coefficient of interest is virtually unaffected. In column 3 we add controls for whether the mother tongues of mother and fathers (inferred from their

[20]The same results hold when we look at fathers or at the whole sample.

[21]This discussion is rather informal given that, even if T_m^m and θ_s are assumed to be positively correlated, the sign of the bias when the school fixed effect is left out depends in general on the pattern of variances and covariances with all the other controls. Moreover, another reason why we would expect a larger role of parental inputs from the specification with host country fixed effects is a milder attenuation bias from measurement error.

countries of origin) are the same of the mother tongue spoken in the host country, while in column 4 we add a measure of linguistic distance from Spolaore and Wacziarg (2015). In both cases the coefficient on T^m remains positive and significant, and does not change much in size.

3.5 Decomposition

For our decomposition we introduce a slightly more general model, which allows parental inputs supplied by both mothers and fathers to differ across countries. Suppose that the test score in year t of child i, studying (and born) in school s (and in country c), whose mother was born in country m and father in country f is given by

$$T_{ist}^{mf} = Parents_{ist}^{mf} + \alpha_{st} + \rho' X_{ist}^{mf} + \varepsilon_{ist}^{mf} \tag{2}$$

where $Parents_{ist}^{mf}$ is the combined effect of all parental inputs and is given by

$$Parents_{ist}^{mf} = \gamma^m + \delta^f + \beta' ParentsEdu_{ist}^{mf} + \eta_{ist}^{mf} \tag{3}$$

with γ^m and δ^f being country specific components capturing a set of average (unobservable) characteristics of mothers and fathers from countries m and f respectively. The parental component of student i includes also the effect of her parents' education, which potentially might influence her performance in school.

Combining (2) and (3) we get our regression of interest:

$$T_{ist}^{mf} = \beta' ParentsEdu_{ist}^{mf} + \gamma^m + \delta^f + \alpha_{st} + \rho' X_{ist}^{mf} + u_{ist}^{mf}$$

This model can be estimated on the sample of students for which both parents are born in a different country from the one where the PISA test is taking place. However, in order to use all the available information in the data and to obtain more precise estimates for the other controls (including the country and school fixed effects), all second generation immigrants and native students can be included in the following specification

$$T_{ist}^{mf} = \beta' ParentsEdu_{ist}^{mf} + \gamma^m + \delta^f + \theta^m NatMoth_{ist}^{mf} + \theta^f NatFath_{ist}^{mf} + \rho' X_{ist}^{mf} + \alpha_{st} + u_{ist}^{mf} \tag{4}$$

where $NatMoth_{ist}^{mf}$ and $NatFath_{ist}^{mf}$ are dummies identifying native parents (mothers and fathers, respectively). Notice that we allow the "native advantage" to be country specific for both mothers and fathers: this is done as otherwise this kind of variation would be absorbed by the country of origin fixed effects, which, in that case, would not be identified only out of second generation immigrants.

The object whose variation we are ultimately interested in decomposing is the average score of native students in country c, which is given by

$$T_c^c = \alpha + Parents^c + \overline{\alpha}_c + \rho' \overline{X}_c \tag{5}$$

where $Parents^c = \gamma^c + \delta^c + \beta' \overline{ParentsEdu}^c$, $\overline{\alpha}_c$ is a weighted average of the school fixed effects in country c and $\overline{X}_c$ and $\overline{ParentsEdu}^c$ are within country c averages.[22] Equation (5) makes our decomposition explicit: our objective is to evaluate the importance of $Parents^c$ to account for the variation of T_c^c across countries.

In order to do that, we estimate our country c specific parental component from

$$\widehat{Parents}^c = \widehat{\gamma}^c + \widehat{\delta}^c + \widehat{\beta}' \overline{ParentsEdu}^c$$

where $\widehat{\gamma}^c, \widehat{\delta}^c$ and $\widehat{\beta}$ are our estimated parameters from (4). Figure 6 plots the parental component obtained from both the school fixed effect and the country fixed effect specifications against the average score of natives (with $Parents_{\text{CHINA}}$ being normalized to 1 in both cases). Not surprisingly, the estimated $Parents^c$ is larger (in absolute terms) for countries that perform better in the PISA test, which means that our parental component does account for some of the cross-country variation (as opposed to masking an even larger dispersion) of average test scores. Consistently with our discussion in Section 8, the dispersion in $Parents^c$ is larger under the country fixed effect specification, which ignores the within country variation in school quality and is not subject to the selection bias arising from assortative matching between parents and schools. Table 12 displays $Parents^c$ for all countries for which we have at least 100 second generation immigrants in our sample and therefore a reasonably precise

[22] The constant now absorbs the "native advantage" terms and an average of the waves fixed effects.

estimate of the corresponding fixed effect.

With our estimate at hand, we can compute the share of the total cross-country variance of T_c^c accounted by $Parents^c$, simply as

$$V_{Parents} = \frac{Var(\widehat{Parents^c})}{Var(T_c^c)}$$

Moreover, for every country (or group of countries) c we can calculate the share of the gap in average test score accounted by the parental component with respect to a given benchmark b as

$$S_{Parents}(c,b) = \frac{\widehat{Parents^b} - \widehat{Parents^c}}{T_b^b - T_c^c}$$

Finally, to gauge the relative contribution of parental education and country specific intercepts, we also compute equivalent statistics for the country specific intercepts only[23],

$$V_{FE} = \frac{Var(\widehat{\gamma}^c + \widehat{\delta}^c)}{Var(T_c^c)}$$

$$S_{FE}(c,b) = \frac{(\widehat{\gamma}^b + \widehat{\delta}^b) - (\widehat{\gamma}^c + \widehat{\delta}^c)}{T_b^b - T_c^c}$$

Tables 11 and 12 show the results of these calculations, both under the country fixed effect and the school fixed effect specifications. For the pairwise comparisons, we focus on the gap between East Asia and the other group of countries as defined in Section 3.2 for Table 11, and on the gap between China and the other countries for Table 12.

The first panel of Table 11 shows that $Parents^c$ accounts for between 12% and 26% of the total variation across countries. However, the role of the parental component is substantially larger when considering the gaps between East Asia and some of the other groups. Particularly striking are the results for Southern Europe and Middle East/North Africa, which as shown in Table 1 display large gaps with respect to the best performing countries, more than a third of which is potentially explained by differences in terms of parental inputs. On the other hand, it is interesting to notice the relatively smaller role that parental inputs play for Latin American countries, whose poor performance in standardized test has been object of recent study (Hanushek and Woessmann, 2012b). For both the variance and pairwise decompositions, the country of origin fixed effects account for the bulk of the contribution

[23] That is, excluding country differences in average parental education. This is to understand how much of the variation is related to parents' observable characteristics nd how much to unobservable characteristics.

of the parental component, suggesting that cross-country differences in parents' education do not play an important role.[24]

Table 12 shows the corresponding results for individual countries. On average, between 17% and 52% of China's out-performance can be accounted by parental inputs. Once again, differences in the estimated country-specific intercepts drive most of the parental component contribution.

3.6 Mechanism

In this section we attempt to open the black box of the parental inputs whose importance was quantified above. What makes parents from high PISA countries more "effective" in terms of the school performance of their children? While answering this question precisely is difficult, we attempt to shed some light on this by proceeding in three steps. First, we distinguish between two alternative interpretations on the source of differences in parental inputs, one based on an intergenerational effect of parental education and another based on a cultural transmission mechanism. Then we turn to the Time Use data to see whether immigrant parents from high PISA countries and their children differ in some observable practices that might help us to explain their better performance at school. Finally, we test whether measures relative to countries' of origin economic development, culture and religion can explain our correlation of interest.

3.6.1 Interactions

The results shown in the previous sections can be rationalized by two conceptually distinct interpretations. One possibility is that the outstanding performance of second generation immigrants from high PISA countries is a by-product of the fact that their parents received an education of higher quality in their country of origin. While conceptually this would still imply that these students have an advantage in terms of parental inputs, the source of this advantage would be the school system itself, creating a powerful intergenerational multiplier effect of educational quality. In other words, while our decomposition would still be valid in

[24] For several geographic regions, and notably for the US, $S_{FE}(c, EA)$ is considerably larger than $S_{Parents}(c, EA)$, since in those regions parents are on average more educated than their East Asian counterparts.

an accounting sense, different parental inputs would only be a proximate cause of differences in test scores, with the underlying fundamental force being the school system instead.

An alternative explanation is that there is some fixed cultural trait which parents transmit to their children, which is unrelated to the quality of parental education and might have its roots in factors deeply entrenched with a country's history and culture. This would mean that this component is likely to be quite persistent over time, and improving a country's educational system might not do much in raising the average test scores if this cultural aspect does not change as well.

An useful way to discriminate between these two views is to explore the heterogeneity of the importance of the country specific parental component with respect to parental education and years since migration. If the intergenerational transmission of educational quality is important, we would expect the correlation with the PISA score in the parents' country of origin to be particularly strong for immigrant parents with higher education in their home country[25], which have been more exposed to the educational system.[26] At the extreme, parents with no education could not transmit the quality of their home country's educational system at all. On the other hand, if the underlying mechanism is a fixed cultural trait we would expect that parents that have been abroad for a longer time might be culturally more integrated in their host country and would have at least in part converged to its cultural norms.[27] In that case, the correlation with the average test score in the country of origin would be weaker for parents that have emigrated many years ago. Moreover, there is some evidence that highly educated immigrants have an easier time integrating in their host country[28]: if this is the case, under the "cultural" interpretation, parents' years of schooling (acquired either in the home or in the host country), would also alleviate the correlation between their children performance and the average score in their country of origin.

To summarize, we have some testable implications to discriminate between the two mecha-

[25]It is actually unclear whether only years of schooling in the home country should matter, giver that there could be dynamic complementarities in the human capital accumulation process that make the impact of an additional year of schooling in the host country stronger for parents that have spent the initial part of their educational career in higher quality schools. Moreover, it is possible that parents emigrating from high PISA countries would go to better schools once in the host country. Since we actually find a negative interaction term, this issue is mostly inconsequential for our purposes.

[26]This line of reasoning is similar to the one in Schoellman (2012), even though here it is applied to returns to parental education for school outcomes of their children.

[27]See Giavazzi et al. (2014) for evidence on the speed of convergence of different cultural traits.

[28]For example, there is widespread evidence that more educated migrants have a higher propensity to intermarry with natives (see Schoen and Wooldredge (1989); Sandefur and McKinnell (1986); Lichter and Qian (2001); Meng and Gregory (2005); Chiswick and Houseworth (2011)), which is an important indicator of integration in the host country.

nisms. The intergenerational transmission of educational quality mechanism would imply a positive interaction term between parents' years of schooling acquired in the home country and the average score of natives in the same country. The fixed cultural trait mechanism would instead predict a negative interaction between the average test score and parents' years since migration, as well as with parents' years of schooling.

We now turn to the US Census data to put these predictions to empirical scrutiny. We once again restrict attention to the results relative to second generation immigrants on the mother side in the main text.[29] We construct a measure of mothers' years of schooling both in their home and in their host countries from information on year of immigration and age at the end of education (imputed from the educational level). Year of immigration is available only as a categorical variable, identifying intervals of approximately 5 years. We therefore impute the exact year of arrival in the US according to two alternative criteria: either we assign to everybody the middle year of their interval (for our first specification below), or we impute the last year in that interval in order to identify immigrants that most likely had their whole education in their home country (for our second specification).

Table 13 shows our main results. We start by adding to the baseline specification in column 1 an interaction term between T^m and mother's years of schooling, finding a negative and significant coefficient. When we break down years of schooling between those acquired in the US and those acquired in country m (columns 3 and 4), we find that the interaction term is negative in both cases, with coefficients of similar magnitudes. These results are inconsistent with the presence of a strong intergenerational effect of educational quality.

A similar message emerges from the study of heterogeneity with respect to years since migration. According to the results in column 5, the correlation between T^m and the school performance of children is weaker for mothers that have emigrated many years ago. Column 6 shows that this result (as well as the results on education previously discussed) is unaffected by the inclusion of controls for age at migration, which has also been shown to be important for the assimilation of immigrants (Bleakley and Chin, 2010; Nielsen and Schindler Rangvid, 2012).

A possible concern is that our imperfect mapping from the information available from the Census and years of schooling accumulated in country m and in the US might confound our results, since the predictions of the "educational quality" mechanism would hold unambiguously only for education acquired in the mother's home country (see the discussion in

[29]The results for the rest of the sample are similar and available upon request.

footnote 25). To alleviate this concern, Table 14 shows results for a subsample of mothers that we are more confident completed their whole education in their country of origin, since we imputed their year of immigration using the most restrictive criterion discussed above. We can see that the interaction between T^m and mother's years of schooling is still negative and significant, as well as the one between T^m and years since migration.

Overall, our results seem supportive of the fixed cultural trait interpretation, given that our correlation of interest is attenuated by both parental education[30] and integration in the host country. To offer further visual evidence for the first fact, Figure 3 and 4 plot, for the PISA and US Census data respectively, the country-specific intercept and coefficient on mother's years of schooling from a regression of our outcome of interest on these variables and the usual controls, with the sample always restricted to second generation immigrants on the mother side. In both cases, the correlation between school performance and test scores in the mother's country of origin is mainly driven by the intercept, and not by different returns to education. This pattern is different from the one documented by Schoellman (2012) for wages of immigrants. Schoellman (2012) shows that returns to education (and not the intercept) of US immigrants are positively related to GDP per capita and PISA scores in the country of origin and interprets this as evidence in favour of the fact that school quality varies greatly across countries. Our results show that, while differences in school quality might be important for labor market outcomes of immigrants, they do not seem to account for the differential school performance of their children. What matters in this case are fixed cultural traits incorporated in the country specific intercepts.

3.6.2 Time Use

In this section we investigate whether the way in which immigrant parents from high PISA countries and their children spend their time might help us explaining our main results.

Table 16 starts by looking at parents. Columns 1 to 3 refer to total child care, while columns 4 to 6 break down the time spent with children in the educational, recreative and basic categories described in Section 3.2. Across all specifications and time use categories, interviewed parents from high PISA countries stand out for spending more time with their children. The result is robust to the inclusion of state fixed effects and several controls on demographic characteritstics of both parents and children. Since time use variables are

[30] We verified that this pattern holds also for the PISA data. These results are not shown and are available upon request.

measured in minutes and refer to a single day, from column 3 it emerges that an increase of one standard deviation in the PISA score in a parent's country of origin corresponds to a higher investment of approximately one hour per week in total child care. This extra hour is quite evenly spread across the three time use subcategories.

Table 18 examines instead time use habits of children. Here we restrict the sample to students between 15 and 18 years old, which are enrolled in full time education; moreover, as usual we restrict attention to second generation immigrants on the mother side in the main text.Our dependent variable is time spent studying, as a proxy for one the main inputs in the educational process. While the sample size is quite small, we find a strong and significant correlation with the average test score of the mother country of birth. According to column 3, the most complete specification, an increase of a standard deviation in T^m is associated with more than 4 extra hours of studying per week.

The results in this section seem to indicate that second generation immigrants do differ in terms of observable practices as a function of the country of origin of their parents. In absence of a credible estimate of the effectiveness of parental child care and children's study time, it is of course difficult to establish where these differences might be driving the results found in the previous sections.

3.6.3 Country Level Characteristics

We now present results from specification (1), augmented by a series of controls at the mother's country of origin level. The main objective of this analysis is to verify that the estimate of our coefficient of interest does not pick up variation across different country-level characteristics, that might also plausibly affect second generation immigrants' school performance.

Table 19 includes controls related to the level of economic development and to the quality of the education system in country m. In particular, we include log GDP per capita, the percentage of native mothers with at least some tertiary education, average years of education of natives between 20 and 30 years old in 1990 and the primary school pupil to teacher ratio in country m. None of these controls significantly affects the coefficient on T^m[31], and, with the exception of the share of skilled of native mothers, they do not seem to affect the performance

[31]Regressions with other controls of school quality are available under request. The main result is unaffected across all specifications we have experimented with.

of second generation immigrants either.

Table 20 includes controls related to religion and cultural values. Columns 2 controls for the religious composition on the population in 1970, while column 3 controls for some answers to various World Value Survey questions that should capture attitudes towards education and hard work. In both cases the coefficient of interest is not greatly affected, even though some of the controls introduced seem to have explanatory power on the performance of second generation immigrants.

3.7 Conclusions

While PISA scores are often taken as a metric to compare the quality of different educational systems, this is not the whole story. In this paper we argue that an important share of the cross-country variation in test scores is driven by differences in broadly defined parental inputs. We arrive to this conclusion by comparing the PISA performance of second generation immigrants, which are born and educated in the same country and school, but have parents of different nationalities. While we cannot entirely rule out the possibility that some unobservable characteristic plays a role in driving our results, we show through various checks that this is unlikely to be the case, and that any residual pattern of differential selection would probably go against finding our result.

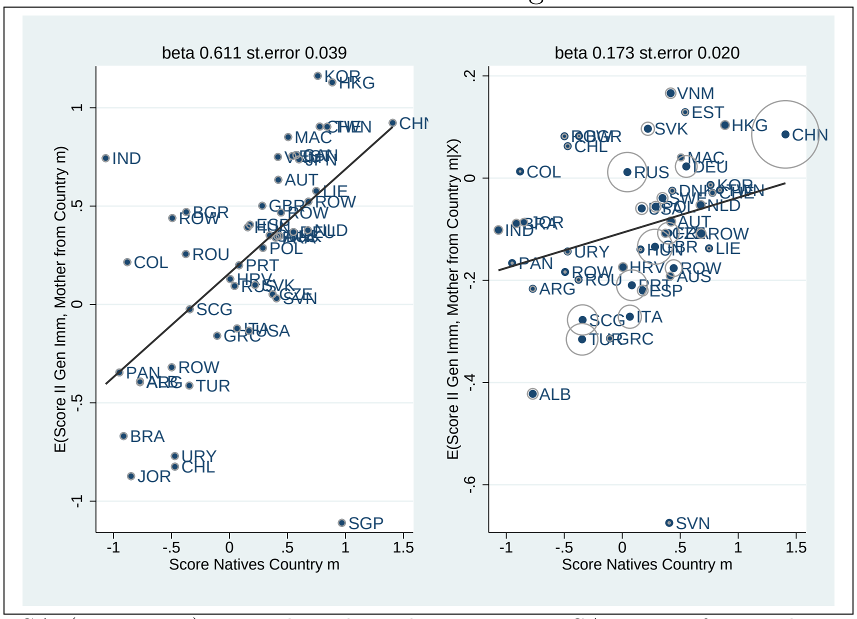
We also provide evidence against the possibility that the superior performance of second generation immigrants from high PISA countries is due to an intergenerational spillover of the high quality education received by their parents. There seems to be some deep cultural factor that makes parents born in some specific countries invest more in their children's education.

Our paper leaves open important avenues for future research. While our main contribution is to identify and quantify an important parental quality component in average PISA scores, our evidence on the specific mechanisms behind this component is only suggestive. We believe it would be important to deepen our understanding of what East Asian parents do differently from Southern European ones (to pick a stark comparison) when raising their children, and whether these differences are optimal responses to the economic environments they are placed in, or maybe the by-product of differences in preferences shaped by the historical experiences of the two regions.

Moreover, our results could be viewed as a sign of caution for policy-makers aiming to raise their students' performance in standardized tests. Since cross-country gaps seem to go beyond differences in school quality, it is unclear to what extent various policies can be effective to this end, given that the cultural factors that lead parents to invest more or less in children's education might be deeply entrenched and persistent over time.

Figures

Figure 1: Performance of Second Generation Immigrants and Natives from Country m



Source: PISA (2003-2012). Panel a plots the average PISA score of second generation immigrants whose mother is from country m (y-axis) with respect to the average PISA score of natives in country m (x-axis). Panel b corrects the average PISA scores of second generation immigrants by differences in observable characteristics (as explained in Footnote 17). The size of the circles in panel b represents the sample size.

Figure 2: Different Types of Selection

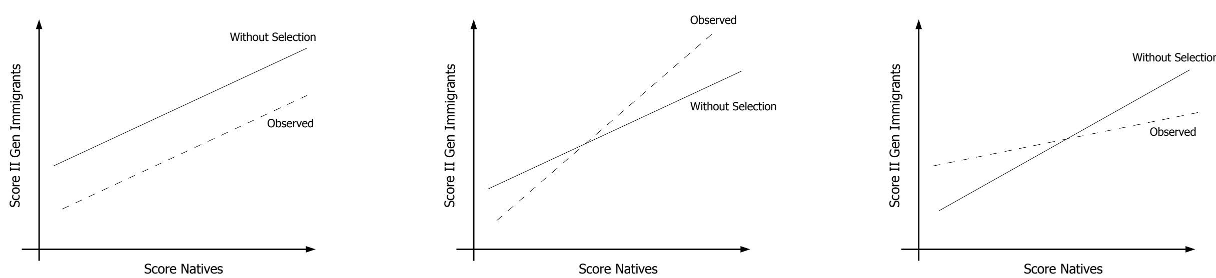
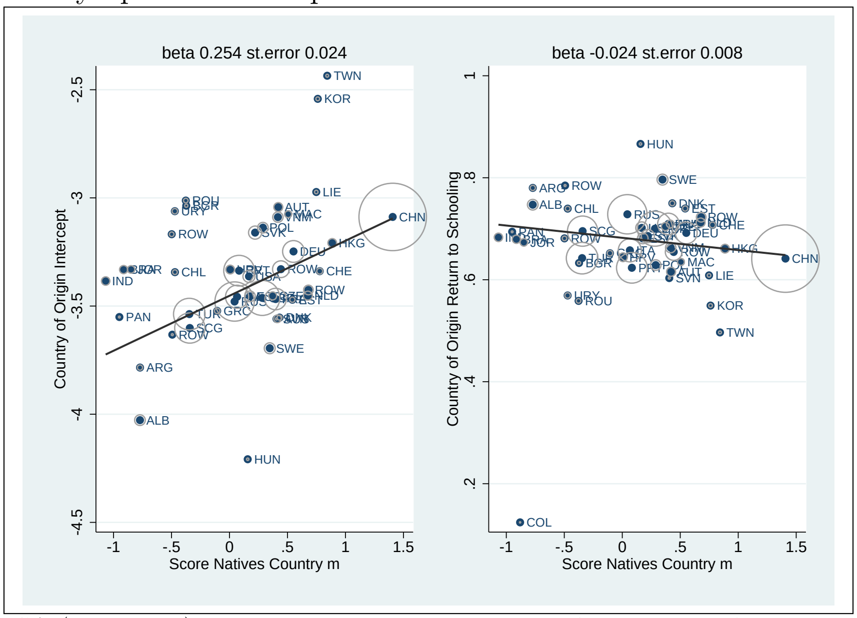
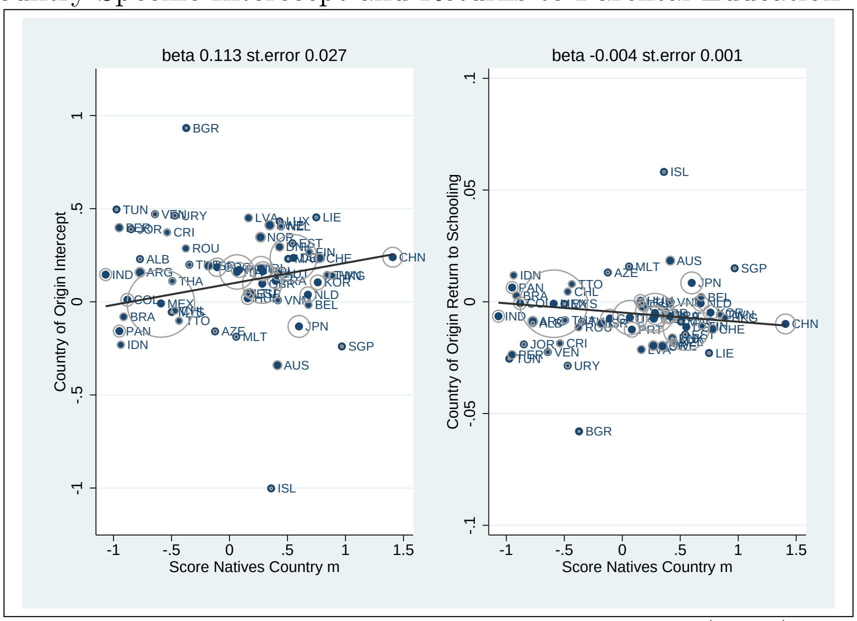


Figure 3: Country Specific Intercept and Returns to Parental Education - PISA Data



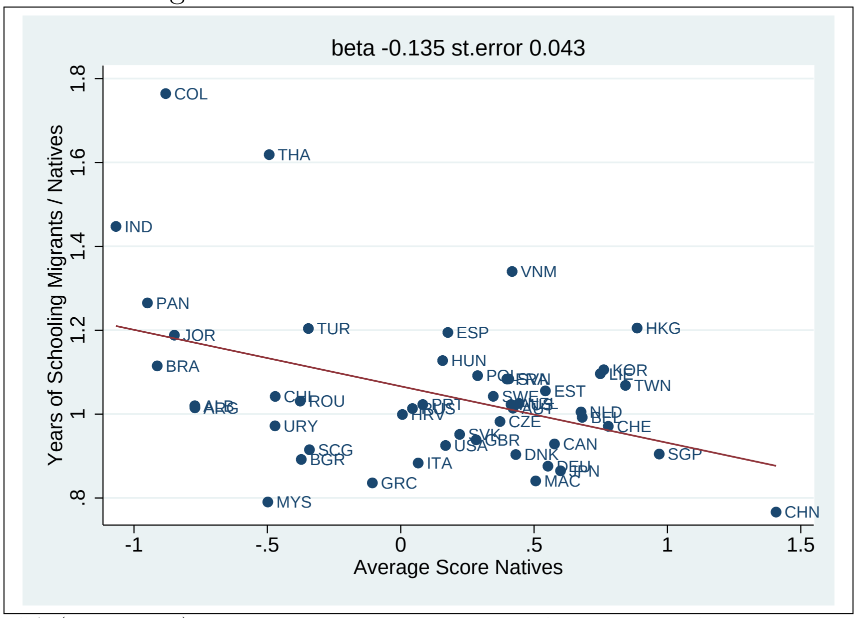
Source: PISA (2003-2012). Panel a plots the country specific intercepts θ_m in equation $T_{ist}^m = \theta^m + \beta^m T^m * EduMoth_{ist}^m + \beta X_{ist}^m + \epsilon_{ist}^m$ (y-axis) with respect to the average PISA score of natives in country m (x-axis). Panel b plots the country specific returns to education β_m with respect to the average PISA score of natives in country m. The size of the circles represents the sample size.

Figure 4: Country Specific Interecept and Returns to Parental Education - US Census



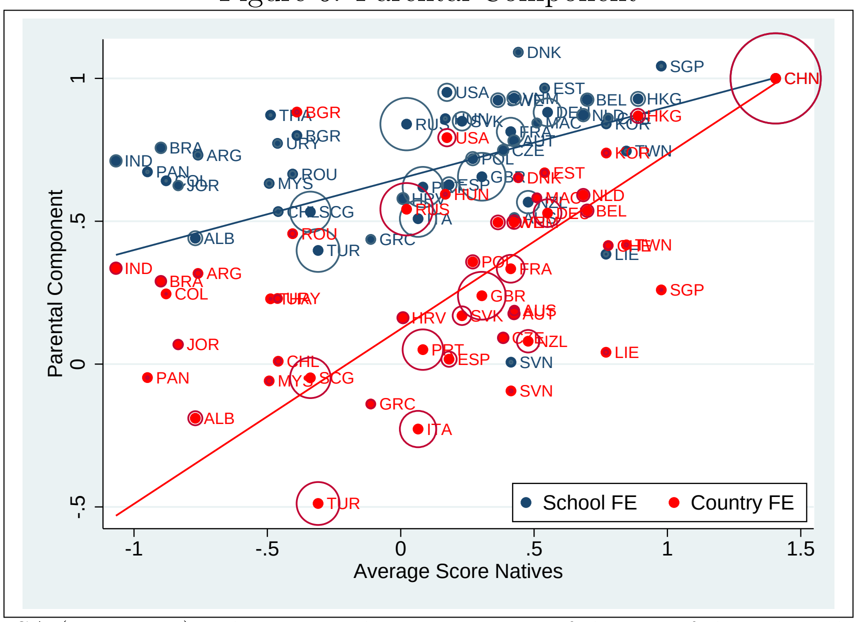
Source: US Census. Panel a plots the country specific intercepts θ_m (y-axis) with respect to the average PISA score of natives in country m (x-axis). Panel b plots the country specific returns to education β_m with respect to the average PISA score of natives in country m. The size of the circles represents the sample size.

Figure 5: Selection on Parental Education



Source: PISA (2003-2012). The Figure plots a measure of the degree of selection into migration in country m (the average years of schooling of emigrant parents over the average years of schooling of remaining parents) (y-axis) with respect to the average PISA score of natives in country m (x-axis).

Figure 6: Parental Component



Source: PISA (2003-2012). The Figure plots an estimate of the size of the parental component ($Parents_c$ in the decomposition described in Section 3.5, normalized such that it takes value 1 for China) (y-axis) with respect to the average PISA score of natives in country m (x-axis).

Tables

Table 1: Average PISA Scores across Regions

	Math	Reading	Science	# Countries
East Asia	0.85	0.58	0.69	7
EU North	0.49	0.43	0.44	15
Oceania	0.43	0.49	0.50	2
US	0.17	0.34	0.30	1
EU South	0.06	0.08	0.10	5
EU East	-0.16	-0.27	-0.15	19
Other Asia	-0.52	-0.51	-0.47	5
Latin America	-0.69	-0.52	-0.59	11
Middle East/NA	-0.71	-0.59	-0.60	7

The Table shows the simple average of the scores obtained in countries belonging to each region, across all available waves. Scores are standardized to have mean 0 and standard deviation 1.

Table 2: Summary statistics - PISA (Mothers)

Variable	Mean	Std. Dev.	Min	Max
Score	.1944	.9961	-3.179	3.362
T^m	.2225	.6063	-1.068	1.407
Mother Sec Edu	.5154	.4998	0	1
Mother Ter Edu	.3001	.4583	0	1
Father Sec Edu	.5142	.4998	0	1
Father Ter Edu	.3351	.472	0	1
Immigrant Mother	1	0	1	1
Immigrant Father	.6486	.4774	0	1
# Obs in Host Country c	1115.0256	2004.6515	10	9759
# Obs from Country m	887.4694	2301.3304	2	14905
# Host Countries for Mothers from Country m	3.3469	3.0452	1	13
Observations		43486		

The Table shows descriptive statistics for second generation immigrants on the mother's side. Sample includes only cases where both parents report a country of origin and the country of origin of the mother runs a PISA test on natives. Scores are from the math test and are standardized to have mean 0 and standard deviation 1 (in the whole sample). Observations weighted accordind to the provided sample weights.

Table 3: Second Generation Immigrants by Country of Origin - PISA Data

	Mothers			Fathers		
Country of Origin	Number	# Host Countries	Top Host Country	Number	# Host Countries	Top Host Country
Albania	360	5	Greece (200)	332	5	Greece (173)
Argentina	83	2	Uruguay (82)	80	1	Uruguay (80)
Australia	169	2	New Zealand (168)	133	2	New Zealand (132)
Austria	260	2	Switzerland (192)	188	2	Switzerland (144)
Belgium	276	3	Luxembourg (261)	258	2	Luxembourg (238)
Brazil	214	4	Uruguay (92)	192	4	Uruguay (86)
Bulgaria	34	1	Turkey (34)	17	1	Turkey (17)
Canada	2	1	Ireland (2)	2	1	Ireland (2)
Chile	71	1	Argentina (71)	57	1	Argentina (57)
China	14905	13	Macao (9570)	14224	11	Macao (8788)
Colombia	5	1	Costa Rica (5)	6	1	Costa Rica (6)
Croatia	229	3	Serbia-Mont. (121)	195	3	Serbia-Mont. (93)
Czech Republic	216	2	Slovakia (206)	187	2	Slovakia (177)
Denmark	82	2	Norway (81)	103	1	Norway (103)
Estonia	84	1	Finland (84)	56	1	Finland (56)
France	1398	7	Switzerland (650)	1181	7	Switzerland (469)
Germany	1470	9	Switzerland (658)	1175	9	Switzerland (478)
Greece	94	2	Australia (70)	144	2	Australia (118)
Hong Kong	248	2	Macao (174)	451	3	Macao (363)
Hungary	19	2	Austria (17)	17	2	Austria (13)
India	234	4	Australia (201)	235	4	Australia (197)
Italy	1630	9	Switzerland (1061)	2805	9	Switzerland (1844)
Jordan	184	1	Qatar (184)	145	1	Qatar (145)
Liechtenstein	40	1	Switzerland (40)	28	1	Switzerland (28)
Macao	149	1	Hong Kong (149)	132	1	Hong Kong (132)
Malaysia	67	4	Australia (54)	57	4	Australia (46)
Netherlands	242	5	Belgium (208)	290	5	Belgium (211)
New Zealand	859	1	Australia (859)	859	1	Australia (859)
Panama	11	1	Costa Rica (11)	16	1	Costa Rica (16)
Poland	313	3	Germany (237)	246	3	Germany (196)
Portugal	2824	4	Luxembourg (1906)	2687	5	Luxembourg (1865)
Romania	62	2	Austria (60)	67	3	Austria (53)
Russia	4770	13	Estonia (1391)	4643	13	Estonia (1390)
Serbia-Mont.	2814	9	Switzerland (1637)	2860	9	Switzerland (1649)
Singapore	9	1	Indonesia (9)	10	2	Indonesia (9)
Slovakia	554	2	Czech Republic (549)	657	2	Czech Republic (652)
Slovenia	15	2	Austria (11)	18	2	Austria (11)
South Korea	48	2	Australia (33)	49	2	Australia (36)
Spain	354	5	Switzerland (336)	432	4	Switzerland (412)
Sweden	383	2	Finland (239)	296	2	Finland (182)
Switzerland	114	1	Liechtenstein (114)	97	1	Liechtenstein (97)
Taiwan	27	1	Hong Kong (27)	10	2	Hong Kong (6)
Thailand	14	1	Finland (14)	2	1	Finland (2)
Turkey	2864	8	Denmark (621)	3134	8	Switzerland (632)
UK	3820	5	Australia (2316)	3975	5	Australia (2500)
United States	457	5	Mexico (228)	586	5	Mexico (360)
Uruguay	88	1	Argentina (88)	82	1	Argentina (82)
Vietnam	327	4	Australia (260)	324	3	Australia (251)
Average	906.1	3.4		911.3	3.4	

The Table shows summary statistics on second generation immigrants on the mother and father side from each country of origin in the PISA sample included in our decomposition exercise, across all available waves. *# Host Countries* is the number of different host countries in which second generation immigrants of a given nationality are observed. *Top Host Country* is the host country where the highest number (reported in brackets) of second generation immigrants of a given nationality are observed.

Table 4: Second Generation Immigrants by Host Country - PISA Data

	Mothers			Fathers		
Host Country	Number	# Countries of Origin	Top Country of Origin (in PISA)	Number	# Countries of Origin	Top Country of Origin (in PISA)
Argentina	631	6	Uruguay (88)	585	6	Uruguay (82)
Australia	9022	17	UK (2316)	9394	17	UK (2500)
Austria	1979	15	Turkey (487)	1965	15	Turkey (519)
Belgium	3126	7	Turkey (434)	3524	7	Turkey (492)
Costa Rica	460	3	Panama (11)	537	3	Panama (16)
Croatia	2160	4	Serbia-Mont. (363)	1948	4	Serbia-Mont. (348)
Czech Republic	780	6	Slovakia (549)	1014	6	Slovakia (652)
Denmark	2712	6	Turkey (621)	2814	6	Turkey (625)
Estonia	1708	2	Russia (1391)	1839	2	Russia (1390)
Finland	1103	10	Sweden (239)	1266	10	Sweden (182)
Georgia	97	2	Russia (69)	76	2	Russia (51)
Germany	1429	10	Turkey (512)	1515	10	Turkey (559)
Greece	1270	3	Russia (214)	760	3	Albania (173)
Hong Kong	5447	4	China (4758)	5296	4	China (4938)
Indonesia	44	5	Singapore (9)	55	5	Singapore (9)
Ireland	1173	17	UK (946)	1043	15	UK (814)
Israel	2321	5	Russia (606)	2474	5	Russia (596)
Kazakhstan	1174	2	Russia (982)	1117	2	Russia (918)
Kyrgyzstan	480	2	Russia (106)	297	2	Russia (106)
Latvia	2295	4	Russia (967)	2593	4	Russia (1107)
Liechtenstein	330	11	Switzerland (114)	281	11	Switzerland (97)
Luxembourg	4448	10	Portugal (1906)	4540	10	Portugal (1865)
Macao	10202	5	China (9570)	9654	7	China (8788)
Mauritius	84	4	China (11)	57	4	China (8)
Mexico	1085	4	United States (228)	1398	4	United States (360)
Moldova	203	3	Russia (68)	192	4	Russia (59)
Netherlands	1741	16	Turkey (203)	1832	16	Turkey (228)
New Zealand	1989	8	UK (528)	2144	8	UK (620)
Norway	1145	3	Sweden (144)	1149	3	Sweden (114)
Portugal	1576	5	Brazil (61)	1353	5	Brazil (64)
Qatar	5908	4	Jordan (184)	5159	4	Jordan (145)
Serbia-Mont.	2333	4	Croatia (121)	1782	4	Croatia (93)
Slovakia	593	3	Czech Republic (206)	583	3	Czech Republic (177)
Slovenia	1841	3	Italy (8)	1880	3	Italy (10)
South Korea	29	5	China (11)	-	-	-
Switzerland	8453	11	Serbia-Mont. (1637)	8320	11	Italy (1844)
Turkey	229	5	Germany (67)	190	5	Germany (33)
UK	2199	7	China (25)	2380	7	China (26)
Uruguay	313	4	Brazil (92)	338	4	Brazil (86)
Average	2156.7	6.3		2137.1	6.2	

The Table shows summary statistics on second generation immigrants on the mother and father side observed in each country in the PISA sample, across all available waves. *# Countries of Origin* is the number of different countries of origin of second generation immigrants in a given host country. *Top Country of Origin (in PISA)* is the country of origin from which the highest number of second generation immigrants in a given host country are observed (number reported in brackets).

Table 5: Main results-PISA

	Dependent Variable: Math Test Score					
	[1]	[2]	[3]	[4]	[5]	[6]
			All			No Asia
T^m	0.676***	0.547***	0.261***	0.255***	0.241***	0.317***
	(0.022)	(0.044)	(0.033)	(0.034)	(0.03)	(0.078)
Female	-0.129***	-0.136***	-0.212***	-0.209***	-0.208***	-0.181***
	(0.027)	(0.024)	(0.019)	(0.019)	(0.019)	(0.027)
Moth Sec Edu				0.008	0.003	0.084
				(0.029)	(0.029)	(0.061)
Moth Ter Edu				0.035	0.027	0.136**
				(0.037)	(0.037)	(0.059)
Fath Sec Edu				0.032	0.033	0.083
				(0.034)	(0.034)	(0.069)
Fath Ter Edu				0.103***	0.103***	0.144**
				(0.040)	(0.040)	(0.069)
N	43486	43486	43486	43486	43486	24408
# Country m	49	49	49	49	49	36
R Squared	0.165	0.248	0.665	0.666	0.666	0.636
Host Country f E	No	Yes	Yes	Yes	Yes	Yes
School FE	No	No	Yes	Yes	Yes	Yes
Continent m FE	No	No	No	No	Yes	Yes

The Table shows results for second generation immigrants on the mother's side. The sample includes only cases where both parents report a country of origin and the country of origin of the mother runs a PISA test on natives. T^m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept, wave fixed effect and a dummy for father immigrant status. Observations weighted according to the provided sample weights. Robust standard errors clustered by mother country of origin. All coefficients and standard errors are estimated according to the "Unbiased Shortcut" procedure (PISA Technical Report, 2009), using the replicate weights provided by PISA. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 6: Summary statistics - US Census (Mothers)

Variable	Mean	Std. Dev.	Min	Max
No Grade Repeated	.8083	.3937	0	1
T^m	.1269	.5382	-1.474	1.407
Mother Sec Edu	.4831	.4997	0	1
Mother Ter Edu	.2062	.4046	0	1
Father Sec Edu	.3915	.4881	0	1
Father Ter Edu	.3347	.4719	0	1
Mother Immigrant	1	0	1	1
Father Immigrant	.4594	.4984	0	1
Yrs Since Migr Mother	20.06	8.749	2	57
Female	.4854	.4998	0	1
Student Age	11.35	2.287	8	15
# Obs from Country m	874.7705	2132.5378	1	13813
Observations		53361		

The Table shows descriptive statistics for second generation immigrants on the mother's side. The sample includes only cases where both parents report a country of origin and the country of origin of the mother runs a PISA test on natives. *No Grade Repeated* is a dummy taking value 1 for students attending a grade larger or equal to the mode of the corresponding year of birth, quarter of birth, state and year cell. *Score Country m* is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother. Observations weighted accordind to the provided sample weights.

Table 7: Main results-US CENSUS

Variables	Dependent variable: 1=never repeated a grade					
	[1]	[2]	[3]	[4]	[5]	[6]
			All			No Asia
T^m	0.057***	0.056***	0.032***	0.029***	0.016***	0.030**
	(0.015)	(0.015)	(0.008)	(0.009)	(0.006)	(0.012)
Female	0.067***	0.067***	0.067***	0.067***	0.067***	0.071***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother Sec Edu			0.054***	0.051***	0.050***	0.049***
			(0.013)	(0.012)	(0.012)	(0.014)
Mother Ter Edu			0.068***	0.064***	0.063***	0.058***
			(0.010)	(0.010)	(0.009)	(0.010)
Father Sec Edu			0.041***	0.041***	0.038***	0.045***
			(0.012)	(0.011)	(0.011)	(0.010)
Father Ter Edu			0.072***	0.073***	0.068***	0.075***
			(0.015)	(0.014)	(0.014)	(0.014)
N	53361	53361	53361	53361	53361	47984
# Country m	61	61	61	61	61	45
R Squared	0.07	0.09	0.10	0.10	0.10	0.10
County FE	No	Yes	Yes	Yes	Yes	Yes
Years Since Migr Mother	No	No	No	Yes	Yes	Yes
Continent m FE	No	No	No	No	Yes	No

T^m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept, child age dummies, parents' age, family size, year fixed effect,state fixed effect, (year specific) quarter of birth fixed effect and father immigrant status. Observations weighted accordind to the provided sample weights. Robust standard errors clustered by mother country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 8: Selection

	Mothers		Fathers	
	Years $\text{Edu}_i/\overline{\text{Years Edu}_m}$		Years $\text{Edu}_i/\overline{\text{Years Edu}_f}$	
	[1]	[2]	[3]	[4]
T^m	-0.126*	-0.222**		
	(0.070)	(0.084)		
T^f			-0.108*	-0.178***
			(0.057)	(0.060)
N	45023	45023	44373	44373
R Squared	0.08	0.47	0.09	0.50
Host Country FE	Yes	Yes	Yes	Yes
School FE	No	Yes	No	Yes

The sample includes only second generation immigrants on the mother side for specifications (1) and (2) and on the father side for specifications (3) and (4). $\overline{\text{Years Edu}_m}$ and $\overline{\text{Years Edu}_f}$ are the average years of education of mothers and fathers in the country of birth of the mother and father respectively. Score Country m and Score Country f refer to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother and the father, across all available waves. All specifications control for intercept and wave fixed effect. Robust standard errors clustered by mother country of origin in specifications (1) and (2) and by father country of origin in specifications (3) and (4) . * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 9: Selection into Schools

	Avg Score School	School FE	Academic Admission	% Qual Teachers	Dropout Rate
	[1]	[2]	[3]	[4]	[5]
T^m	0.208**	0.201**	0.052**	0.028**	-0.009**
	(0.084)	(0.079)	(0.022)	(0.011)	(0.004)
Female	0.062***	0.074***	0.014	0.007	-0.002
	(0.020)	(0.019)	(0.010)	(0.007)	(0.002)
Father Sec Edu	0.102***	0.099***	0.061**	-0.021	-0.007
	(0.028)	(0.027)	(0.027)	(0.014)	(0.009)
Father Ter Edu	0.252***	0.238***	0.084***	0.004	-0.008
	(0.040)	(0.037)	(0.028)	(0.016)	(0.010)
Mother Sec Edu	0.157***	0.152***	0.018	0.010	-0.036
	(0.025)	(0.024)	(0.015)	(0.007)	(0.028)
Mother Ter Edu	0.248***	0.234***	0.024	0.009	-0.037
	(0.049)	(0.047)	(0.024)	(0.015)	(0.025)
N	42944	42895	43486	32356	10184
# Country m	49	49	49	48	41
R Squared	0.34	0.35	0.17	0.41	0.06
Host Country FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

The sample includes only second generation immigrants on the mother side. Score Country m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept and father immigrant status. Robust standard errors clustered by mother country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 10: Language

	Dependent Variable: Math Test Score			
	[1]	[2]	[3]	[4]
T^m	0.236***	0.232***	0.236***	0.181***
	(0.029)	(0.035)	(0.035)	(0.029)
Female	-0.206***	-0.207***	-0.206***	-0.171***
	(0.025)	(0.019)	(0.020)	(0.020)
Father Sec Edu	0.027	0.026	0.027	0.131***
	(0.046)	(0.033)	(0.033)	(0.049)
Father Ter Edu	0.084*	0.080**	0.084**	0.202***
	(0.049)	(0.039)	(0.040)	(0.054)
Mother Sec Edu	0.008	-0.001	0.008	0.006
	(0.035)	(0.028)	(0.028)	(0.044)
Mother Ter Edu	0.061	0.052	0.061*	0.070
	(0.050)	(0.037)	(0.037)	(0.046)
Foreign Language at Home		-0.094***		
		(0.031)		
Mother Same Native Lang			0.004	
			(0.034)	
Father Same Native Lang			-0.036	
			(0.043)	
Mother Linguistic Distance				0.120**
				(0.049)
Father Linguistic Distance				0.010
				(0.053)
N	40915	40915	40915	21346
# Country m	49	49	49	42
R Squared	0.664	0.664	0.664	0.582
Host Country FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

The Table shows results for second generation immigrants on the mother's side. Sample includes only cases where both parents report a country of origin and the country of origin of the mother runs a PISA test on natives. T^m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept, wave fixed effect and a dummy for father immigrant stautus. Observations weighted according to the provided sample weights. Robust standard errors clustered by mother country of origin. All coefficients and standard errors are estimated according to the "Unbiased Shortcut" procedure (PISA Technical Report, 2009), using the replicate weights provided by PISA. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 11: Decomposition Results

	$V_{Parents}$ (%)		V_{FE} (%)	
	School FE	Host Country FE	School FE	Host Country FE
	11.9	26.0	11.6	23.6
	$S_{Parents}(c, EA)$ (%)		$S_{FE}(c, EA)$ (%)	
	School FE	Host Country FE	School FE	Host Country FE
East Asia vs				
EU North	17.9	55.6	19.1	60.2
Oceania	37.5	83.3	37.7	84.6
US	2.7	16.4	4.7	22.5
EU South	32.7	79.1	30.4	70.1
EU East	18.9	46.9	19.9	50.8
Other Asia	11.0	35.7	8.1	25.8
Latin America	10.1	30.5	9.0	26.1
Middle East/NA	33.8	81.6	31.0	68.8

The Table shows results from the decomposition of the cross-country variation in the PISA Math Score. The first panel shows the share of the cross-country variance accounted by the whole parental component ($V_{Parents}$) and by the country specific intercept (V_{FE}), while the second panel shows the share of the score gap between East Asia and each group accounted by the whole parental component ($S_{Parents}(c, EA)$) and by the country specific intercept ($S_{FE}(c, EA)$). For each group, $S_{Parents}(c, EA)$ and $S_{FE}(c, EA)$ are computed using weighted averages of the corresponding parental components and test scores for each belonging country, where the weights are given by the (weighted) number of immigrant parents from each country available in the sample.

Table 12: Parental Component by Country

		$Parents_c$		$S_{Parents}(c, \text{CHINA})$ (%)		$S_{FE}(c, \text{CHINA})$ (%)	
Country	PISA Score	School FE	Country FE	School FE	Country FE	School FE	Country FE
China	1.41	1	1	-	-	-	-
Hong Kong	.89	.94	.84	11.9	30.83	6.48	9.93
Switzerland	.78	.95	.53	8.12	74.37	10.29	81.31
Belgium	.7	1	.55	.66	63.41	4.67	75.3
Netherlands	.68	.93	.72	10.08	38.72	12.24	45.36
Germany	.55	.92	.6	8.82	47.17	9.69	50.56
Estonia	.54	.97	.67	3.75	38.41	5.29	43.72
Macao	.51	.94	.65	6.61	39.28	2.35	22.28
New Zealand	.48	.66	.21	36.13	84.79	36.07	85.18
Denmark	.44	1.09	.68	-9.66	33.08	-7.05	41.56
Vietnam	.43	.93	.53	6.94	47.97	1.23	28.16
Australia	.43	.6	.33	40.72	68.27	41.81	71.88
Austria	.42	.87	.27	12.76	73.98	13.35	76.75
France	.41	.81	.36	18.74	64.64	19.32	67.75
Czechia	.38	.75	.37	24.46	61.66	23.86	59.5
Sweden	.36	.93	.52	7.18	46.07	10.26	55.57
United Kingdom	.3	.75	.4	22.51	54.53	23.32	58.05
Poland	.27	.72	.38	24.67	54.74	22.91	50.26
Slovakia	.23	.85	.23	12.64	65.08	11.69	62.47
Spain	.18	.63	.08	30.22	75.03	29.05	69.43
United States	.17	.96	.8	2.87	16.36	4.8	22.52
Portugal	.08	.62	.05	28.81	71.62	25.09	56.48
Italy	.07	.51	-.18	36.66	87.72	35.7	84.85
Russia	.02	.84	.53	11.48	33.64	13.27	39.96
Croatia	.01	.58	.18	30.09	58.64	30.65	60.47
Greece	-.11	.44	-.08	37.07	71.31	36.87	70.25
Turkey	-.31	.4	-.45	35.07	84.41	32.05	70.66
Serbia & Montenegro	-.34	.53	-.03	26.89	59.23	27.66	61.89
Romania	-.41	.66	.45	18.61	30.26	19.01	31.8
Uruguay	-.46	.87	.42	6.8	31.1	5.69	26.4
Chile	-.46	.63	.15	19.61	45.38	18.9	42.51
Malaysia	-.49	.63	.01	19.24	52.12	18.24	49.17
Argentina	-.76	.83	.43	7.7	26.1	7.08	23.12
Albania	-.77	.44	-.18	25.71	54.41	25.15	52.44
Jordan	-.83	.72	.23	12.34	34.36	12.26	33.75
Brazil	-.9	.77	.35	9.77	28.25	8.27	22.48
India	-1.07	.71	.37	11.63	25.67	9.99	19.71
Average	0.10	0.76	0.33	17.16	52.02	16.88	50.65

The Table includes only countries with at least 100 second generation immigrants in the sample. $Parents_c$ is the estimated parental component, normalized such that $Parents_{\text{CHINA}} = 1$.

Table 13: Interactions - US CENSUS

Variables	Dependent variable: 1=never repeated a grade					
	[1]	[2]	[3]	[4]	[5]	[6]
T^m	0.032***	0.112***	0.032***	0.111***	0.155***	0.175***
	(0.008)	(0.024)	(0.008)	(0.026)	(0.034)	(0.040)
Female	0.066***	0.066***	0.066***	0.066***	0.066***	0.067***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Yrs Schooling Father	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Yrs Schooling Mother	0.006***	0.006***				
	(0.001)	(0.001)				
$T^m *$ Yrs Schooling Mother		-0.005***				
		(0.001)				
Yrs Schooling Moth in US			0.007***	0.007***	0.002*	0.005***
			(0.001)	(0.001)	(0.001)	(0.001)
Yrs Schooling Moth in m			0.006***	0.006***	0.007***	0.006***
			(0.001)	(0.001)	(0.001)	(0.001)
$T^m *$ Yrs Schooling Mother in US				-0.006***	-0.002**	-0.003**
				(0.001)	(0.001)	(0.001)
$T^m *$ Yrs Schooling Mother in m				-0.005***	-0.005***	-0.005***
				(0.001)	(0.001)	(0.001)
Yrs Since Migr Mother					0.004***	0.004***
					(0.001)	(0.001)
$T^m *$ Yrs Since Migr Mother					-0.002**	-0.003**
					(0.001)	(0.001)
Age Migration Moth						0.006***
						(0.002)
(Age Migration Moth)^2						-0.000***
						(0.000)
$T^m *$ Age Migration Mother						-0.001
						(0.001)
N	52853	52853	52853	52853	52853	52853
# Country m	61	61	61	61	61	61
R Squared	0.10	0.10	0.10	0.10	0.10	0.10
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

T^m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept, child age dummies, parents' age, family size, year fixed effect,state fixed effect, (year specific) quarter of birth fixed effect and father immigrant status. Observations weighted accordind to the provided sample weights. Robust standard errors clustered by mother country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 14: Interactions - US CENSUS (Mothers Entirely Educated in Home Country)

Variables	Dependent variable: 1=never repeated a grade			
	[1]	[2]	[3]	[4]
T^m	0.042***	0.115***	0.169***	0.185***
	(0.010)	(0.022)	(0.037)	(0.039)
Female	0.068***	0.068***	0.068***	0.068***
	(0.005)	(0.005)	(0.005)	(0.005)
Yrs Schooling Father	0.007***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Yrs Schooling Mother	0.004***	0.004***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
$T^m *$ Yrs Schooling Mother		-0.005***	-0.005***	-0.005***
		(0.001)	(0.001)	(0.001)
Yrs Since Migr Mother			0.005***	0.005***
			(0.001)	(0.001)
$T^m *$ Yrs Since Migr Mother			-0.003**	-0.003**
			(0.001)	(0.001)
Age Migration Moth				0.007**
				(0.003)
(Age Migration Moth)^2				-0.000**
				(0.000)
$T^m *$ Age Migration Moth				-0.001
				(0.001)
N	30118	30118	30118	30118
# Country m	61	61	61	61
R Squared	0.12	0.12	0.12	0.12
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

The sample includes only cases where the mother was entirely educated in her home country. T^m refers to the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. All specifications control for intercept, child age dummies, parents' age, family size, year fixed effect,state fixed effect, (year specific) quarter of birth fixed effect and father immigrant status. Observations weighted according to the provided sample weights. Robust standard errors clustered by mother country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 15: Summary statistics - Time Use Survey (Parents)

Variable	Mean	Std. Dev.	Min	Max
Total Child Care	86.76	122.3	0	1055
Educational Child Care	9.809	30.29	0	420
Recreational Child Care	21.11	59.76	0	639
Basic Child Care	55.85	90.96	0	1055
T^p	-.3779	.5812	-1.068	1.407
Yrs Since Migration	15.7	10.63	1	58
Mother	.4878	.4999	0	1
Age Parent	37.72	8.456	18	80
Age Spouse	37.95	8.862	16	80
Parent Sec Edu	.2397	.4269	0	1
Parent Ter Edu	.3762	.4845	0	1
Spouse Sec Edu	.2277	.4194	0	1
Spouse Ter Edu	.3932	.4885	0	1
Number of Children	2.038	.9657	1	7
Avg Age Children	8.133	5	0	18
Number of Male Children	1.038	.8499	0	5
# Obs from Country p	91.2105	351.2035	2	2624
Observations		5199		

The Table shows descriptive statistics for interviewed immigrant parents in the Time Use Survey. *Total* refers to the total time spent in child care activities, while *Educational*, *Recreational* and *Basic* refer to the time use categories as defined in the text. *Score Country p* is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the interviewed parent. Observations weighted according to the provided sample weights.

Table 16: Time Use of Parents

	Total	Total	Total	Educational	Recreational	Basic
	[1]	[2]	[3]	[4]	[5]	[6]
T^p	15.673**	12.863***	8.432**	1.446**	3.345**	3.641
	(7.209)	(4.661)	(3.445)	(0.720)	(1.445)	(2.366)
Mother			67.800***	7.396***	3.477	56.927***
			(4.029)	(0.667)	(3.524)	(2.490)
Parent Sec Edu			-5.409	3.113***	-4.550	-3.972
			(5.403)	(0.619)	(2.863)	(2.433)
Parent Ter Edu			10.286**	4.335***	-2.597	8.548***
			(5.086)	(1.077)	(3.227)	(2.482)
Spouse Sec Edu			3.116	-2.795***	7.935***	-2.024
			(3.507)	(0.517)	(2.770)	(1.274)
Spouse Ter Edu			12.009***	0.709	6.812*	4.488**
			(2.927)	(1.727)	(3.558)	(2.220)
Age Parent			0.494	0.134**	0.142	0.219
			(0.480)	(0.063)	(0.450)	(0.145)
Age Spouse			0.347*	0.194**	-0.062	0.215
			(0.206)	(0.084)	(0.262)	(0.197)
Number of Children			16.905***	2.817***	1.074	13.014***
			(2.048)	(0.957)	(0.865)	(1.626)
Avg Age Children			-9.135***	-0.353***	-3.386***	-5.397***
			(1.317)	(0.119)	(0.540)	(0.774)
Number of Male Children			-0.652	1.157***	-2.035**	0.226
			(1.744)	(0.417)	(0.794)	(1.204)
Yrs Since Migration			-0.018	-0.153***	-0.070	0.206*
			(0.161)	(0.032)	(0.098)	(0.122)
N	5199	5199	5199	5199	5199	5199
# Country p	57	57	57	57	57	57
R Squared	0.01	0.03	0.23	0.06	0.10	0.21
State FE	No	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes	Yes

The sample includes only immigrant parents of children with at most 18 years. *Total* refers to the total time spent in child care activities, while *Educational*, *Recreational* and *Basic* refer to the time use categories as defined in the text. Score Country p is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the interviewed parent, across all available waves. Additional controls in specifications (3) to (6) are dummies for retired, full time students and disabled parents. Robust standard errors clustered by parent country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 17: Summary statistics - Time Use Survey (Students)

Variable	Mean	Std. Dev.	Min	Max
Study Time	57.86	88	0	525
T^m	-.3441	.5669	-1.068	1.407
Native Father	.1949	.3966	0	1
Female	.422	.4944	0	1
Mother Sec Edu	.2153	.4115	0	1
Mother Ter Edu	.3841	.487	0	1
Father Sec Edu	.1928	.3949	0	1
Father Ter Edu	.3801	.486	0	1
Age	16.39	1.039	15	18
# Obs from Country m	10.5610	38.2590	1	248
Observations		433		

The Table shows descriptive statistics for second generation immigrants on the mother side. *Score Country m* is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother. Observations weighted according to the provided sample weights.

Table 18: Time Use of Students

Variables	Dependent Variable: Study Time		
	[1]	[2]	[3]
T^m	42.154***	41.223***	32.539**
	(13.964)	(12.971)	(12.348)
Native Father	-24.708	-18.913	-30.387
	(20.407)	(23.861)	(24.880)
Female			-4.227
			(8.876)
Mother Sec Edu			-7.142
			(11.836)
Mother Ter Edu			26.791**
			(12.959)
Father Sec Edu			21.461
			(15.656)
Father Ter Edu			16.860
			(16.197)
Age			1.429
			(4.576)
N	433	433	433
# Country m	41	41	41
R Squared	0.07	0.17	0.21
State FE	No	Yes	Yes
Year FE	No	Yes	Yes

The sample includes only second generation immigrants on the mother side that are full time students and at most 18 years old. Score Country m is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. Robust standard errors clustered by mother country of origin. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 19: Country of Origin Characteristics - Economic and Educational Development

Variables	Dependent variable: test score math					
	[1]	[2]	[3]	[4]	[5]	[6]
T^m	0.265***	0.289***	0.204***	0.236***	0.254***	0.244***
	(0.032)	(0.041)	(0.035)	(0.039)	(0.036)	(0.039)
Female	-0.208***	-0.207***	-0.205***	-0.207***	-0.208***	-0.203***
	(0.026)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Father Sec Edu	0.031	0.029	0.03	0.031	0.031	0.025
	(0.047)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Father Ter Edu	0.108**	0.107***	0.106***	0.108***	0.108***	0.098**
	(0.052)	(0.040)	(0.040)	(0.040)	(0.040)	(0.039)
Mother Sec Edu	-0.004	-0.004	-0.018	-0.009	-0.006	-0.02
	(0.036)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Mather Ter Edu	0.026	0.027	0.006	0.019	0.023	0.003
	(0.050)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Log GDP		-0.055**				-0.064**
		(0.027)				(0.030)
% Skilled moth in m			0.207***			0.487***
			(0.078)			(0.119)
Avg Years Edu in m				0.011		-0.006
				(0.009)		(0.011)
Pri Pupil/Teacher in m					-0.001	0.009***
					(0.002)	(0.002)
N	42730	42730	42730	42730	42730	42730
# Country m	44	44	44	44	44	44
R Squared	0.672	0.672	0.673	0.672	0.672	0.674
Host Country FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

The sample includes only second generation immigrants on the mother side. Score Country m is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. Observations weighted according to the provided sample weights. Robust standard errors clustered by mother country of origin. All coefficients and standard errors are estimated according to the "Unbiased Shortcut" procedure (PISA Technical Report, 2009), using the replicate weights provided by PISA. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 20: Country of Origin Characteristics - Religion and Culture

Variables	Dependent variable: test score math			
	[1]	[2]	[3]	[4]
T^m	0.265***	0.261***	0.265***	0.238***
	(0.032)	(0.068)	(0.039)	(0.052)
Female	-0.208***	-0.205***	-0.208***	-0.206***
	(0.026)	(0.019)	(0.019)	(0.019)
Father Sec Edu	0.031	0.031	0.031	0.032
	(0.047)	(0.034)	(0.034)	(0.034)
Father Ter Edu	0.108**	0.104***	0.108***	0.108***
	(0.052)	(0.040)	(0.040)	(0.040)
Mother Sec Edu	-0.004	-0.014	-0.004	-0.008
	(0.036)	(0.028)	(0.028)	(0.029)
Mother Ter Edu	0.026	0.008	0.026	0.020
	(0.050)	(0.038)	(0.037)	(0.038)
% Catholic in m	-0.116			
	(0.112)			
% Protestant in m	0.062			
	(0.133)			
% Other Christian Rel in m	0.055			
	(0.299)			
% Orthodox in m	-0.081			
	(0.190)			
% Jews in m	4.225			
	(4.033)			
% Muslim in m	-0.168			
	(0.126)			
% Jews in m	0.547***			
	(0.183)			
% Buddhist in m	0.04			
	(0.457)			
% Eastern Religions in m	0.591*			
	(0.329)			
% Other Religion in m	-0.008			
	(0.383)			
Leisure Not Important in Life				-0.378***
				(0.110)
Child Quality: Hard Work				-0.055
				(0.118)
Child Quality: Obedience				0.152
				(0.243)
Locus of Control				0.08**
				(0.038)
N	42730	42730	42730	42730
# Country m	49	49	46	46
R Squared	0.672	0.674	0.672	0.673
Host Country FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

The sample includes only second generation immigrants on the mother side. T^m is the average math PISA score of natives (standardized to have mean 0 and standard deviation 1) in the country of birth of the mother, across all available waves. Observations weighted according to the provided sample weights. Robust standard errors clustered by mother country of origin. All coefficients and standard errors are estimated according to the "Unbiased Shortcut" procedure (PISA Technical Report, 2009), using the replicate weights provided by PISA. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

3.8 Appendix

3.8.1 PISA

Given that the individual host countries have great flexibility in choosing how to report parents' countries of birth, some aggregation is necessary to get a set of countries/regions consistently defined over time. In particular, we make the following adjustments: we code *Yugoslavia* and similar labels as Serbia and Montenegro, *USSR* and similar labels as Russia, *Albania or Kosovo* as Albania, *France or Belgium* as France, *Germany or Austria* as Germany, *China (including Hong Kong)* as China. We drop all observations with inconsistent or missing information on students' or parents' countries of birth.

Parents' educational attainment is reported according to the ISCED 1997 classification system. We group levels 0 and 1 into *primary* education, levels 2, 3 and 4 into *secondary* education and levels 5 and 6 into *tertiary* education.

We construct some measures of school quality from the available information in the Student and School Questionnaires. In particular, *Avg Score School* is the average PISA score of the native students in the same school, *Score FE* is the estimated school fixed effect in a regression of the PISA score on gender and parental education dummies (once again limiting the sample to native students), *Academic Admission* is a dummy that takes value 1 whenever schools report that student's record of academic performance is either *always* or *sometimes* considered for admissions, *% Qual Teachers* is the share of current teachers with at least the ISCED 5A level of education and *Dropout Rate* is the share of students who leave the school without having obtained the corresponding diploma.

3.8.2 Time Use

The time use categories are constructed as follows. *Basic Child Care* includes Physical care for household children, Organization & planning for household children, Looking after household children (as a primary activity) , Attending household children's events, Waiting for/with household children, Picking up/dropping off household children, Talking with/listening to household children, Caring for & helping household children (n.e.c.), Providing medical care to household children, Obtaining medical care for household children, Waiting associated with household children's health, Activities related to household child's health (n.e.c.), Travel related to caring for and helping household children. *Educational Child Care* includes Reading to/with hh children, Homework (household children), Meetings and school conferences (household children), Home schooling of household children, Waiting associated with household children's education, Activities related to household child's education (n.e.c.). *Recreational Child Care* includes Playing with household children (not sports), Arts and crafts with household children, Playing sports with household children. *Study Time* includes Research/homework for class for degree, certification, or licensure,Research/homework for class for personal interest, Waiting associated with research/homework, Research/homework (n.e.c.).

Bibliography

Daniel Aaronson, Lisa Barrow, and William Sander. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1):95–135, 2007.

Mark Aguiar and Erik Hurst. Measuring Trends in Leisure: The Allocation of Time over Five Decades. *Quarterly Journal of Economics*, 122(3):969–1006, 2007.

Alberto Alesina and Paola Giuliano. The Power of the Family. *Journal of Economic Growth*, 15(2):93–125, 2010.

Alberto Alesina, Paola Giuliano, and Nathan Nunn. On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, 101(3):499–503, 2013.

Joseph G. Altonji. The Effects of High School Curriculum on Education and Labor Market Outcomes. *Journal of Human Resources*, 30(3):409–438, 1995.

Joseph G Altonji, Erica Blom, Costas Meghir, et al. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1):185–223, 2012.

Massimo Anelli and Giovanni Peri. Peers Composition Effects in the Short and in the Long Run: College Major, College Performance and Income. Working Paper 9119, IZA, 2015.

Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004.

Peter Arcidiacono, V. Joseph Hotz, and Songman Kang. Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics*, 166(1):3–16, 2012.

Peter Arcidiacono, Esteban Aucejo, and V. Joseph Hotz. University Differences in the Graduation of Minorities in STEM Fields: Evidence from California. *American Economic Review*, forthcoming.

Robert D. Atkinson and Merrilea Mayo. Refueling the U.S. innovation economy: Fresh Approaches to Science, Technology, Engineering and Mathematics (STEM) Education. Working papers, The Information Technology and Innovation Foundation, 2010.

David H. Autor. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, 21(1):1–42, 2003.

David H. Autor and David Dorn. The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–97, 2013.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Incentives for managers and inequality among workers: evidence from a firm-level experiment. *Quarterly Journal of Economics*, 122(2):729–773, 2007.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *Review of Economic Studies*, 77(2):417–458, 2010.

Robert J. Barro and Jong Wha Lee. A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104(C):184–198, 2013.

Robert J. Barro and Rachel M. McCleary. Religion and Economic Growth across Countries. *American Sociological Review*, 68(5):760–781.

Magali Beffy, Denis Fougere, and Arnaud Maurel. Choosing the Field of Study in Postsecondary Education: Eo expected Earnings Matter? *Review of Economics and Statistics*, 94(1):334–347, 2012.

Stefanie Behncke. How Do Shocks to Non-Cognitive Skills Affect Test Scores? *Annals of Economics and Statistics*, pages 155–173, 2012.

Roland Benabou and Jean Tirole. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3):489–520, 2003.

Eric Bettinger and Long Bridget Terry. Does Cheaper Mean Better? The Impact of using Adjunt Instructors on Student Outcomes. *Review of Economics and Statistics*, 92(3): 598–613, 2010.

Julian R. Betts and Heather Rose. The Effect of High School Courses on Earnings. *Review of Economic Studies*, 86(2):497–513, 2004.

Hoyt Bleakley and Aimee Chin. Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants. *American Economic Journal: Applied Economics*, 2(1):165–92, 2010.

Lex Borghans, Huub Meijers, and Bas Ter Weel. The Role Of Noncognitive Skills In Explaining Cognitive Test Scores. *Economic Inquiry*, 46(1):2–12, 2008.

George J Borjas. Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics*, 107(1):123–50, 1992.

Michela Braga, Marco Paccagnella, and Michele Pellizzari. Evaluating Students Evaluations of Professors. *Economics of Education Review*, 41:71–88, 2014.

Michela Braga, Marco Paccagnella, and Michele Pellizzari. The Academic and Labor Market Returns of University Professors. *Journal of Labor Economics*, Forthcoming.

James A. Brickley and Jerold L. Zimmerman. Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate Finance*, 7(4): 367–396, 2001.

Alexander Brüggen and Frank Moers. The role of financial incentives and social incentives in multi-task settings. *Journal of Management Accounting Research*, 19(1):25–50, 2007.

Giorgio Brunello and Martin Schlotter. Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems. IZA Discussion Papers 5743, Institute for the Study of Labor, 2011.

Stephen V. Cameron and James J. Heckma. The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, 109(3):455–499, 2001.

Pedro Carneiro, Claire Crawford, and Alissa Goodman. The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes. CEE Discussion Papers 0092, Centre for the Economics of Education, LSE, 2007.

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

Daniele Checchi, Gianni De Fraja, and Stefano Verzillo. Publish or Perish? Incentives and Careers in Italian Academia. IZA Discussion Papers 8345, Institute for the Study of Labor, 2014.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9): 2593–2632, 2014.

Barry Chiswick and Christina Houseworth. Ethnic intermarriage among immigrants: human capital and assortative mating. *Review of Economics of the Household*, 9(2):149–180, 2011.

Chua, Amy. *Battle Hymn of the Tiger Mother*. The Penguin Press, 2011.

Kalena Cortes, Joshua Goodman, and Takako Nomi. Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra. *Journal of Human Resources*, 50:108–158, 2015.

Giacomo De Giorgi, Michele Pellizzari, and Silvia Redaelli. Identification of Social Interactions through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics*, 2(2):241–75, 2010.

Shaun Dougherty, Joshua Goodman, Darryl Hill, Erica Litke, and Lindsay Page. Middle School Math Acceleration and Equitable Access to 8th Grade Algebra: Evidence from the Wake County Public School System. *Educational Evaluation and Policy Analysis*, 37: 80S–101S, 2015.

Jaap Dronkers and Manon de Heus. Immigrants' children scientific performance in a double comparative design: The influence of origin, destination, and community. CReAM Discussion Paper Series 1213, Centre for Research and Analysis of Migration, Department of Economics, University College London, 2012.

Etienne Dumont, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics*, 27(6):1436–1450, 2008.

Christian Dustmann, Stephen Machin, and Uta Schonberg. Ethnicity and educational achievement in compulsory schooling. *Economic Journal*, 120(546):272–297, 2010.

Christian Dustmann, Tommaso Frattini, and Gianandrea Lanzara. Educational Achievement of Second-generation Immigrants: an International Comparison. *Economic Policy*, 27(69): 143–185, 2012.

Glenn Ellison and Ashley Swanson. Heterogeneity in High Math Achievement Across Schools: Evidence from the American Mathematics Competitions. NBER Working Papers 18277, National Bureau of Economic Research, Inc, 2012.

European Commission. The European Commission: EUROPE 2020 Strategy. Technical report, 2010.

Ernst Fehr and Urs Fischbacher. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478):C1–C33.

Susan Feng Lu. Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes. *Journal of Economics Management Strategy*, 21(3):673–705, 2012.

Raquel Fernandez. Does culture matter? In *Handbook of Social Economics*. Elsevier, 2011.

Raquel Fernandez and Alessandra Fogli. Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, 1(1):146–77, 2009.

David N. Figlio and Lawrence W. Kenny. Individual Teacher Incentives and Student Performance. *Journal of Public Economics*, 91(5-6):901–914, 2007.

David N. Figlio, Morton O. Schapiro, and Kevin B. Soter. Are Tenure Track Professors Better Teachers? *Review of Economics and Statistics*, 97(4):715–724, 2015.

Hans Fricke, Jeffrey Grogger, and Andreas Steinmayr. Does Exposure to Economics Bring New Majors to the Field? Evidence from a natural Experiment. Working Papers 21130, National Bureau of Economic Research, 2015.

Roland G. Fryer and Richard T. Holden. Multitasking, Dynamic Complementarities, and Incentives: A Cautionary Tale. Technical report, Cambridge, Massachusetts: Harvard University, 2013.

Francesco Giavazzi, Ivan Petkov, and Fabio Schiantarelli. Culture: Persistence and Evolution. NBER Working Papers 20174, National Bureau of Economic Research, Inc, 2014.

Paola Giuliano. Living arrangements in western europe: Does cultural origin matter? *Journal of the European Economic Association*, 5(5):927–952, 2007.

Claudia Goldin. A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104:1091–1119, 2014.

Joshua Goodman. The Labor of Division: Returns to Compulsory Math Coursework. Faculty research working paper series, Harvard University, John F. Kennedy School of Government, 2012.

Katja Gorlitz and Christina Gravert. The Effects of a High School Curriculum Reform on University Enrollment and the Choice of College Major. Working Papers 8983, IZA, 2015.

Eric A. Hanushek and Ludger Woessmann. Do Better Schools lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17(4): 267–321, 2012a.

Eric A. Hanushek and Ludger Woessmann. Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics*, 99(2):497–512, 2012b.

Justine S. Hastings, Christopher A. Neilson, and Seth D. Zimmerman. Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile. NBER Working Papers 19241, National Bureau of Economic Research, Inc, 2013.

James J. Heckman and Yona Rubinstein. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2):145–149, 2001.

Bengt Holmstrom and Paul Milgrom. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization*, 7(0):24–52, 1991.

Bengt Holmstrom and Paul Milgrom. The Firm as an Incentive System. *American Economic Review*, 84(4):972–91, 1994.

Fuhai Hong, Tanjim Hossain, John A. List, and Migiwa Tanaka. Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories. Working paper, National Bureau of Economic Research, 2013.

Caroline Hoxby. Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Papers 7867, National Bureau of Economic Research, Inc, 2000.

Information Technology and Innovation Foundation. Refueling the U.S. Innovation Economy: Fresh Approaches to STEM Education. Technical report, 2010.

Jun Ishii and Steven G Rivkin. Impediments to the Estimation of Teacher Value Added. *Education Finance and Policy*, 4(4):520–536, 2009.

Brian A. Jacob. Accountability, Incentives and Behavior: the Impact of High-stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6):761–796, 2005.

Estelle James, A. Nabeel, and J. Conaty. College Quality and Future Earnings: Where Should You Send Your Child to College? *American Economic Review*, 79(2):247–52, 1989.

John Jerrim. Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent. *Oxford Review of Education*, 41(3):310–333, 2015.

Ning Jia. Do Stricter High School Math Requirements Raise College STEM Attainment? Working Papers 8983, Mimeo, 2014.

Juanna Schrter Joensen and Helena Skyt Nielsen. Is there a Causal Effect of High School Math on Labor Market Outcomes? *Journal of Human Resources*, 44(1):171–198, 2009.

Juanna Schrter Joensen and Helena Skyt Nielsen. Mathematics and Gender: Heterogeneity in Causes and Consequences. *Economic Journal*, pages n/a–n/a, 2015.

Carlos Munoz Johnson, Ryan and David H. Reiley. The War for the Fare': How Driver Compensation Affects Bus System Performance. *Economic Inquiry*, forthcoming.

Charles I. Jones. Sources of U.S. Economic Growth in a World of Ideas. *American Economic Review*, 92(1):220–2395, 2002.

Thomas J. Kane and Douglas O. Staiger. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. NBER Working Papers 14607, National Bureau of Economic Research, Inc, 2008.

Dennis D. Kimko and Eric A. Hanushek. Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review*, 90(5):1184–1208, 2000.

Lars J. Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of Study, Earnings and Self-selection. *Quarterly Journal of Economics*, 2016.

Cory Koedel and Eric Tyhurst. Math Skills and Labor-Market Outcomes: Evidence from a Resume-Based Field Experiment. Working Papers 1013, Department of Economics, University of Missouri, 2010.

Edward P. Lazear. Performance Pay and Productivity. *American Economic Review*, 90(5): 1346–1361, 2000.

Phillip B Levine and David J Zimmerman. The Benefit of Additional High-School Math and Science Classes for Young Men and Women. *Journal of Business & Economic Statistics*, 13(2):137–49, 1995.

Daniel T. Lichter and Zhenchao Qian. Measuring Marital Assimilation: Intermarriage Among Natives and Immigrants. *Social Science Research*, 30:289–312, 2001.

Wiswall M and B Zafar. Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies*, 2014.

Xin Meng and Robert G. Gregory. Intermarriage and the Economic Assimilation of Immigrants. *Journal of Labor Economics*, 23(1):135–176, 2005.

Jacob A. Mincer. *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc, 1974.

Enrico Moretti. *The New Geography of Jobs*. Houghton Mifflin Harcour, 2012.

Helena Skyt Nielsen and Beatrice Schindler Rangvid. The Impact of Parents' Years since Migration on Children's Academic Achievement. 2012.

OECD. PISA Data Analysis Manual: SPSS. Technical report, 2009.

OECD. PISA 2009 Technical Report. Technical report, 2011.

OECD. PISA 2012 Results: What Students Know and Can Do. Technical report, 2012a.

OECD. PISA 2012 Technical Report. Technical report, 2012b.

Philip Oreopoulos and Marianne E. Page. The Intergenerational Effects of Compulsory Schooling. *Journal of Labor Economics*, 24(4):729–760, 2006.

Ronni Pavan and Josh Kinsler. The Specificity of General Human Capital: Evidence from College Major Choice. *Journal of Labour Economics*, 33(4):933–972, 2015.

Abigail Payne and Joanne Roberts. Government Oversight of Public Universities: Are Centralized Performance Schemes Related to Increased Quantity or Quality? *Review of Economics and Statistics*, 92(1):207–212, 2010.

Giovanni Peri, Kevin Shih, and Chad Sparber. STEM Workers, H1B Visas and Productivity in US Cities. Norface Discussion Paper Series 2013009, Norface Research Programme on Migration, Department of Economics, University College London, 2013.

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

Canice Prendergast. Intrinsic Motivation and Incentives. *American Economic Review*, 98 (2):201–205, 2008.

Andrew Rendall and Michelle Rendall. Math Matters: Education Choices and Wage Inequality. ECON - Working Papers 160, Department of Economics - University of Zurich, 2014.

Steven G Rivkin, Eric A Hanushek, and John F Kain. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458, 2005.

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

Jesse Rothstein. Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables. *Education Finance and Policy*, 4(4):537–571, 2009.

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

Gary Sandefur and Trudy McKinnell. American Indian Intermarriage. *Social Science Research*, 15(4), 1986.

Todd Schoellman. Education Quality and Development Accounting. *Review of Economic Studies*, 79(1):388–417, 2012.

Robert Schoen and John Wooldredge. Marriage Choices in North Carolina and Virginia, 1969-1971 and 1979-1981. *Journal of Marriage and the Family*, 51:465–481, 1989.

Marianne E. Page Scott E. Carrell and James E. West. Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics*, 125(3):1101–1144, 2010.

Enrico Spolaore and Romain Wacziarg. Ancestry, Language and Culture. CESifo Working Paper Series 5388, CESifo Group Munich, 2015.

Ralph Stinebrickner and Todd R. Stinebrickner. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *Review of Economic Studies*, 81(1):426–472, 2014.

The President's Council of Advisor on Science and Technology. Engange to Excel: Producing one Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. Technical report, 2012.

UK Government. Science & Innovation Investment Framework 2004-2014. Technical report, Department of Education, HM treasury, Department of Health, Department of Trade and Industry, 2004.

UK HM Treasury and BIS. The Plan For Growth. Technical report, 2010.

John V. Winters. Foreign and Native-Born STEM Graduates and Innovation Intensity in the United States. IZA Discussion Papers 8575, Institute for the Study of Labor, 2014.