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

Marta De Philippis†

September 8, 2015

Abstract

Increasing the number of Science, Technology, Engineering and Math (STEM) graduates at university is considered a key element for long-term productivity, innovation and competitiveness in the global economy. Still, little is known about what actually drives and shapes students' choices. This paper explores how being exposed to more science in high school affects enrollment and persistency in STEM majors, the quality of the university attended and the probability of dropping out of high school, by exploiting 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. I find that taking more science in high school increases the probability of enrolling in a STEM major by 25% and the probability of graduating in these majors by the same amount. Moreover the probability of dropping out from high school decreases for low income students. Finally, the effect is not limited to natural sciences only: students choose to combine more science in high school with more difficult high school and post-high school courses.

*VERY PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE WITHOUT THE AUTHOR'S PERMISSION. I thank Steve Pischke and Esteban Aucejo for very precious guidance, supervision and encouragement. I thank Lorenzo Cappellari, Georg Graetz, Monica Langella, Alan Manning, Barbara Masi, Stephan Maurer, Sandra McNally, Guy Micheals, Michele Pellizzari, Paolo Sestito, Olmo Silva, Alessandro Vecchiato and Giulia Zane and participants to the LSE labour and education work in progress seminars, to the 2015 CEP conference, to the 5th fRDB workshop and to the 6th IWAEE workshop for providing me with very useful comments and information. The views expressed in this article are those of the author alone and do not necessarily reflect the official views of the Bank of Italy.

†London School of Economics, Centre of Economic Performance, Bank of Italy. E-mail address: m.de-philippis@lse.ac.uk

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 a key element for long-term productivity and competitiveness [Winters, 2014, Peri et al., 2013, Moretti, 2012, Atkinson and Mayo, 2010, Jones, 2002]. STEM majors, moreover, not only generate high social returns in terms of countries growth and productivity, but also appear as a very profitable private investment for college graduates. Lifetime earnings of graduates in STEM majors are extremely high [Kirkeboen et al., 2015, 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 2009 the wage gap between the average electrical engineer and someone majored in general education US was almost identical to the wage gap between the average college graduate and the average high school graduate. The differences in earnings among university majors, moreover, 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 returns of scientific and mathematical majors. Moreover, engineering is considered a better investment than Harvard [James et al., 1989]: graduates in STEM fields earn more independently of the quality of the institution they attend [Kirkeboen et al., 2015, Arcidiacono et al., 2013]. 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. STEM occupations are ranked among the first.

However, despite the high social and provate benefits obtained from 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 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 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. I address in particular three questions. First, does offering advanced science courses in high school increase by itself the supply of STEM graduates? Is it just a matter of preparation or does it partly come from changes in self-confidence? Second, who is induced to take advanced science when it is offered? Only students who will benefit from it directly, i.e. that are willing to take scientific majors later on, or also other students who benefit in other, indirect, ways? Third, this paper evaluates whether more

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

[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.

exposure to science at age 14 works for everybody, or the effect is concentrated on some segment of the population. Is it more efficient to offer the science courses in high school to everybody or to target the offer to some subgroups of the population only, like high ability students or low income families? Some papers [Dougherty et al., 2015] find that universal coverage of stronger mathematics courses may not be overall beneficial. Finally this paper evaluates whether my findings are driven by other correlated changes, like different peers or teachers in the advanced science classes. This helps identifying the exact origin of the effect in order to allow policymakers to reproduce the policy in other contexts.

The identification of the effect of 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. I eliminate the selection of different courses within the same school by collapsing the analysis at the school level (in the spirit of Altonji [1995]). I address instead the other layer of selection, i.e. 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 strategies that approach this type of selection problem in two different and complement ways. 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 compares students unexpectedly exposed to the offer of the advanced science course, because their schools start to offer it by the time they have to choose their field courses, with students not exposed to the offer, because their schools do not offer it.[5] One possible limitation, of which I will take extensive care, is that schools decide to offer advanced science exactly based on the specific characteristics of the current cohort in the school. I first show that this is not the case, at least for observable characteristics. Then, I implement my second identification strategy that is not subject to this concern because it uses differences 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 oversubscribed, select students based to home-to-school distance and schools catchment areas vary (unpredictably) over time. The instrument therefore compares students living in the same neighbourhood, who are more or less likely to enroll in schools offering advanced science, because of exogenous changes in schools' catchment areas.

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 30% and of enrolling in STEM majors by about 25%. Offering more science courses at high school moreover not only induces more students to enroll in STEM majors but also it increases

[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, but in a context where 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 and by anticipation of the effect.

the likelihood that they graduate in these majors. This is extremely important, given the huge problem with persistency in this kind of majors [Arcidiacono et al., 2013, Stinebrickner and Stinebrickner, 2014]. Second, I find that the main determinant behind the choice of taking advanced science, when students are exposed to this option, are previous test scores in science. Boys and girls at this stage behave similarly. 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 enrollment therefore increases as a consequence of this policy; not because less girls take advanced science at age 14: the share of girls and boys taking the advanced science class is more or less equal, but because girls, when exposed to more science in high school, are induced to take, yes, more difficult 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 at high schools. I estimate that the policy contributed to increase the share of STEM graduates in the UK by 5% between 2005 and 2010 (out of a total increase by 17%). I estimate moreover that if all schools started offering advanced science to the top 30 students in each cohort the share of STEM graduates would increase by 11%

This paper speaks to the literature aimed at explaining individuals' choices and, in particular, 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 are expected earnings; competencies and preparation; self-confidence; preferences and innate ability [Arcidiacono et al., 2012, Arcidiacono, 2004, Beffy et al., 2012, Stinebrickner and Stinebrickner, 2014, M and Zafar, 2015]. Since preferences and ability are considered to be fixed over time, returns to STEM majors are already very high, and Stinebrickner and Stinebrickner [2014] show that students start university being overconfident about their scientific ability, there is large scope for policies intervening on students' preparation and on primary and secondary school quality.[7] 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. While some studies look at the effects of some school inputs (usually at the university level), like peers [De Giorgi et al., 2010, Anelli and Peri, 2013], 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 a 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.[8]

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

[6] i.e. subjects usually taken by high ability students

[7] Many scholars [Cameron and Heckma, 2001, Moretti, 2012], indeed, attribute the lack of STEM graduates to the low quality of the US school system.

[8] 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.

possible selection of students into schools offering different curricula. 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) sciences 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 classroom heterogeneity and composition.[9] I disentangle the curriculum effect from the peer channel, using as instrument for peers 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 controlling for changes in peers' characteristics.

Third, the compliers in my setting are extremely high ability students with very high willingness to enroll in STEM majors. These are the students of highest interest for policymakers because they are more likely to actually enroll in STEM majors and to make important contributions to scientific and technological fields. Most of the existing empirical works [Goodman, 2012, Kalena et al., 2015] analyze instead policies that affect 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].[10]

The remainder of this paper is organized as follows. In Section 2, I describe the data, the English school system and the advanced science program. Section 3 provides an overview of the main identification strategy. In Section 4, I present the estimated impact of advanced science on post-16 educational outcomes, I check my identifying assumptions and the robustness of my results. Moreover, Section 5 disentangles the mechanisms behind my estimates and Section 6 concludes.

2 Data and Institutional Setting

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 move to Key Stage 1 (KS1), spanning ages 5 and 6, and Key Stage 2 (KS2, age 7 and 11).[11] 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

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

[10] 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.

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

subjects.[12] In particular, out of usually between 10 and 12 qualifications, students typically choose between 4 and 6 subjects.[13] 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, 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[14], and are free to choose any subject. Finally, higher education usually begins at age 19 with a three-year bachelor's degree.

I look at the effect of subjects chosen in KS4 on KS5 and university subject and participation choices.

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 (which is worth 1 KS4 qualification). They can, moreover, choose to take 'double science' (worth 2 qualifications) which leads to more knowledge in all the three subjects or 'triple sciences' (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.

In 2004 the UK Government published a ten-year investment framework for science and innovation [UK Government, 2004]. The framework set out the Government's 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 consider 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 top 40% of students) was established. In particular the government stated that "all pupils achieving at least level 6[16] at Key Stage 3 should be entitled to study triple science at KS4, for example through collaborative arrangements with other schools, Further Education colleges and universities". 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 on high ability students (from 15% to 50%).

There are several, mainly supply driven, reasons why schools may introduce the triple science option in different years. 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

[12] 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 mainly consider GCSE qualifications or GCSE equivalent qualifications.

[13] 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.

[14] 50% of students takes between 3 and 3.5 A level equivalent qualifications.

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

[16] Level 6 or above is equivalent to the top 30% of students.

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.[17] 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.

2.3 Data

I use administrative data on all students in maintained schools in the UK, that follow students from primary school till the end of secondary school. Moreover, I was able to link these data to the universe of all UK university students. In this way, my 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 addresses and postcodes. I obtain instead information on students' attainments in all their Key Stages exams (from KS1 till KS5) from the National Pupil Database (NPD). Moreover there is information on every single subject chosen (and the grade) in KS4 and KS5. Finally, NPD provides a very rich set of data on school characteristics (peer group, type of school, teachers' hirings, school location etc.).

From the NPD dataset I also obtain the information on which courses are offered by different schools. In particular, I follow the official methodology used by the 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. HESA provides me with information on whether pupils progress to university on their major and on the institution they attended. I combined these data to create a dataset following the entire population of five cohorts of English school children. These are 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 2004, 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. 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 when she had to choose her KS4 subjects.
Finally, I impose a set of 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.[18] This leaves me with approximately 530,000

[17] 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.

[18] I checked whether this selection generates any bias and this i snot the case. The results are available upon

students per cohort.

The data I use are a major improvement over previous studies. The very detailed nature of the information I need on subject choices, gives particularly large scope for measurement error problems in survey data and makes it difficult for me to rely on survey data only for my analysis. However, students' administrative data available in other countries do not have all elements I need to run my analysis. Most datasets do not have information on University major choices and outcomes. Other administrative datasets include post high school outcomes as well, but usually refer to small countries, relatively homogeneous in terms of students' background and do not include information on previous test scores. The large amount of observations and the heterogeneity of students backgrounds 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.

3 Empirical Strategy

3.1 A 'nested' selection problem

As stated in the introduction, 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:

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

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 would not provide the right counterfactual ($E(Y_0|D = 0) \neq E(Y_0|D = 1)$). In practice the bias is generated by a 'nested selection' problem, because there 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 follow:

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

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

$$Y^j_{ist} = \beta_1 D_{ist} + \beta_2 X_{ist} + \delta_s + \delta_t + u^j_{ist} \tag{3}$$

where $j = 1, 0$ refers to whether D_{ist}, the usual dummy equal to 1 if student i in high school s, in cohort t takes triple science and 0 otherwise; X_{ist} are school and student controls, δ_s are

request.

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 3, I get:

$$Y_{ist} = \beta_1 D_{ist} + \beta_2 X_{ist} + \delta_s + \delta_t + \underbrace{u^0_{ist} + D_{ist}(u^1_{ist} - u^0_{ist})}_{\epsilon_{ist}} \tag{4}$$

where β_1 is the effect of more science in high school on subsequent subject choices.

Estimating equation 4 by OLS is likely to lead to inconsistent and probably upward biased estimates of β_1, because of selection bias. Adding controls to equation 4 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 level 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 my instrument is at the school level. The error term becomes $\overline{\epsilon}_{st} = \overline{u}^0_{st} + D(\overline{u}^1_{st} - \overline{u}^0_{st})$ and varies only at the school-year level. Most papers [Altonji, 1995], by using school average curriculum as instrument, basically use this approach and therefore address this type of selection only.

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

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 both selected schools not offering triple science at age 11. My treated group is composed by students whose schools unexpectedly started to offer triple science by the time they turned 14.

My first stage equation is:

$$D_{ist} = \gamma_1 z_{st} + \gamma_2 X_{ist} + \zeta_s + \zeta_t + v_{istj} \tag{5}$$

where z_{st} is a dummy equal to one 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; 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.[19]

[19] In principle, I could also use as control group the schools already offering triple science, after having controlled for whether school s was offering triple science when students in cohort t applied, in order to take care of selection into schools. However, trends are not parallel in this case because a curriculum reform, which took place in 2006, affected schools already offering triple science differently from schools not offering triple science.The reform introduced a new KS4 curriculum, more focused on explaining "how science works", and provided greater flexibility to students in terms of which science option to take and when to choose it, by introducing units and modules common to different type of qualifications. This reform, allowing students to more flexibly choose whether to take triple science or not, dramatically increased the number of students taking triple science in English schools.

This strategy mainly relies on two assumptions.

First, the assumption that the information set of both students in the treatment and students in the control groups 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. In Section 4.4, I show that parents/students do not select schools strategically.

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 2.2 I described some supply driven reasons why schools may delay the introduction of triple science. In Section 4.4 I test this second identifying assumption.

3.3 Second Instrument

Still, schools may decide when to offer triple science depending on unobservable characteristics of their current cohort. 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 enroll 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 geographical distance.[20] 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 center is the school and ray is the maximum observed home-to-school distance,[21] 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)[22] where student i used to live at age 9. 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 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:

$$D_{ipt} = \theta_1 z_{pt}^2 + \theta_3 X_{ipt} + \theta_t + \theta_p + v_{ipt} \tag{6}$$

[20] With some exceptions for students with siblings attending the same school or for students with special education needs.

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

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

where z^2_{pt} is my instrument: the share of schools reachable in year t by student i, residing in neighbourhood p, that were offering triple science in year t, when i applied to secondary school. X_{ipt} are neighbourhood and individual controls[23] and θ_t and θ_p are cohort and neighbourhood fixed effects respectively.

This instrument compares students attending schools that offer triple science with students attending schools not offering it. 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 relevant than differences within schools over time.. Section 4.5 addresses this concern.

4 Results

In this section I begin by showing results obtained with my first instrument. I first show the overall effect of taking more science in high school in term of post-16 outcomes. Second, I describe who decides to take triple science, when exposed to the option of taking it, by characterizing compliers. Third, I analyze how the effect is heterogenous, depending on gender, ability and socio-economic status. Finally I check my identifying assumptions and whether my main results are robust to the second identification strategy.

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.[24] The Table proceeds by showing how different specifications estimate the effect of interest. 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^1_{st}), but instead of including school fixed effects only adds some school level controls.[25]. Column 5 shows results from my preferred specification, and estimates equation 4, using the identification strategy whose first stage is described by equation 5. Column 6 includes a school specific trend. Reassuringly, the coefficients of columns 5 and 6 are very similar, suggesting that schools offering triple scienc are not on a different trend. In Column 7 I estimate the specification of equation 4, but I eliminate controls, to check whether my instrument 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 almost 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 (30% over the mean) and 1.5 percentage point more likely to choose a STEM major at university (25% over the mean).

Table 3 shows the coefficients obtained from estimating my main equation on some age 14 (KS4) outcomes, other age 16 (KS5) outcomes and other university outcomes. The top panel

[23] Average primary school grade in math, english, science of students in the neighbourhood, share of girls and FSME, beside the usual individual controls

[24] I adopt throughout the paper a narrow definition of STEM majors, that includes engineering, math, the natural sciences and medicine. The dependent variables 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.

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

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 consequence of the new course offered. Results show that the number of exams taken at age 14 slightly increases.

The second panel refers to outcomes at 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 in the likelihood of choosing science subjects, conditional on continuing. Column 1 tells that the coefficient estimated on STEM majors comes entirely from an increase in the second component, because the first is not affected by the policy. The other columns show whether there is an effect on other subjects. The results show that the effect of studying triple science is not only limited to the pure natural science subjects but it also has spillovers on math and English for instance. If we believe exposure to more science only works because of lack of preparation in high school, then triple science should only have an effect on scientific majors. If, instead, taking triple science also has an impact on other, non scientific but difficult subjects, then it means that a part of the effect might act through changes in self-confidence.

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.[26]. The other columns show the effect on other subjects. Finally the coefficients displayed in the last column of the second panel of Table 3 shows that 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 by more or less the same amount.[27] This is extremely relevant given the large debate that is taking place in the US about the low persistence of students in scientific fields [Arcidiacono et al., 2013, Stinebrickner and Stinebrickner, 2014].

One important aspect, which due to lack of available data is largely underexplored 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 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 15 in the Appendix shows the coefficients and standard errors obtained from estimating equation 4 using each time a different KS4 subject as dependent variable.[28] Tables 16 and 17 report the same type of estimates but they refer 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 later on.Triple science increases the probability of choosing scientific subjects, 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.

[26]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.

[27]The results on university outcomes are estimated on the 2005-2007 sample only, otherwise there is no information on whether the students graduated from university.

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

However, it is difficult to draw conclusions from these coeffients, anecdotical evidence may suggest that a vocational course in music is very different from an advanced course in science at age 14, but to objectively measure the degree of difficulty of each option Table 4 uses a more formal procedure. I define courses along three dimensions: (i) 'high achievers' courses, characterized by a high average primary school grade of students choosing them in out-of-sample academic years; (ii) scientific courses, characterized by a purely scientific course content[29] and (iii) 'female dominated' courses, characterized by a high share of girls attending the courses in previous academic years. Figure 6 describes the data. In particular it shows three scatterplots where for each course is displayed on the x-axis the share of girls enrolled and on the y-axis the average primary school grade of student enrolled. Triple science is by far the most difficult course at KS4, followed by foreign languages, history and geography. For what concern KS5 options, math is the most difficult, followed by physics, chemistry and foreign languages. For university majors, medicine, languages and STEM subjects are very difficult while education, subjects allied to medicine and art are the least difficult subjects. The correlation between the difficulty of each course and the share of girls enrolled in those courses is negative. This is surprising, given that on average girls have higher grades than boys in primary school.

Table 4 shows how much does the average course difficulty at age 18 (KS5) and at university increase as a consequence of taking more science at KS4. 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) difficult subjects, different from the three natural sciences.[30] The results show that taking advanced science at age 14 induces students to choose more difficult subjects later on and the effect is not only driven by pure science courses. Students taking triple science are induced to choose courses at age 16 whose difficulty is about 0.2 standard deviations higher. This 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%). The same is true for university majors.

4.2 Compliers' Characterization

This Section analyzes who decides to take triple science, when the school offers it. This is useful to understand, on one side, how students makes decisions about which subjects to take at age 14, on the other side, whether heterogeneous results are driven by differences in the actual treatment effect or 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 or not triple science is a free decision made by students.[31]

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 increase 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 option, will not be the same for all students: very good students as well as students with very strong preferences towards other

[29] For KS4 and KS5 I define natural science courses, courses in chemistry, biology and physics. For university I define a course as scientific if depending on the number of natural science KS5 courses required to enroll in that major

[30] To obtain these results I multiply the coefficients displayed in Tables 15, 16 and 17 by the numbers displayed in Figure 6 and i sum the series. Standard errors are computed through the Delta method.

[31] 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 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.

subjects may not find beneficial to take triple science. This means that the likelihood of taking triple science will not be the same for everybody: it will depend on students preferences, on students innate ability and on their perceptions towards their abilities.

The first row of Table 5 shows my 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.

The Table 5 then characterizes compliers for the entire population and boys and girls, repsectively. 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.[32] Table 5 displays coefficients from estimating Equation 5, 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 ability in primary school is more than two. Moreover compliers tend to be high income students and 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.

4.3 Heterogeneity

This section evaluates whether strengthening the science curriculum in high school is more effective for certain subgroups of the population. In particular I analyze the heterogeneity of the effect by gender, socio-economic status and previous attainments.

The first panel of Table 6 looks at whether attending more science classes at high school has a different effect depending on students previous science grades (in primary school). In particular the Table looks at the probability of enrolling in STEM majors and of persisting in these studies. It shows that 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 lower ability students are instead less likely to be affected by the policy at all.

The second panel analyzes heterogeneity by socio-economic status (SES).[33] The Tables shows that the effct on KS5 is larger for low SES students. The effect on university outcomes is, instead, more difficult to estimate with enough precision because of the small share of low SES students attending university.

The third panel looks at whether the effect is stronger for boys or girls. The effects are positive for both genders, but the effect on STEM majors looks stronger for boys. Still, girls are more likely to enroll in scientific majors as well, but tend to choose more female-dominated science majors like medicine instead of engineering.

Table 7 shows results obtained in Table 4, splitting the sample by gender. While it is true that girls tend to choose more difficult subjects, they still opt for more female-dominated subjects (like medicine for instance).

[32] First stages in this case do not include any control a part from year and school fixed effects.

[33] 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 hh in the area). Table 6 uses only FSM, but results are consistent with both proxies.

4.4 Checks to the identification strategy

As stated in Section 3, this instrument 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 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 applied) and I estimate the following equation:

$$W_{ist} = \alpha_1 z_{st}^{11} + \alpha_2 z_{st} + \alpha_3 X_{ist} + \xi_s + \xi_t + \eta_{ist} \tag{7}$$

where W_{ist} are some outcomes (like whether student i chooses a STEM major or graduates in it) and some 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 had to choose her school and z_{st} is my usual instrumental variable. The second panel of Table 8 includes also school specific trends. Table 8 shows the coefficient α_1 is not significant for most variables, suggesting that, even if parents and students know when they choose their school whether $S = 1$ or 0, they do not select schools correspondingly. 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 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 [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[34] Column 5 refers to my preferred specification that 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 14 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 starts offering triple science, the trend was parallel to that of all other schools still not offering triple science. To do this, I augment my reduced form regression with leads and lags of the instrument (following Autor [2003]):

$$y_{ist} = \sum_{t=0}^{m} \gamma_{\tau-t} z_{s(\tau-t)} + \gamma_2 \sum_{t=0}^{q} \gamma_{\tau+t} z_{s(\tau+t)} + \zeta_t + \zeta_s + u_{ist} \tag{8}$$

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

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

parallel pre-treatment trends by evaluating whether all $\gamma_{\tau-t}$ are close to 0. Figure 5 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. Figure 5 shows the same results, but using predetermined characteristics: 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. This confirms the results obtained in Table 9 and 8.

Another 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.[35] This would imply that part of the effect I find is purely mechanical. I address this concern in Table 10. Column 1 looks at how the probability of offering science at KS5 evolves over time and whether it corresponds to the cohort when the school starts offering triple science at KS4 as well. The correlation is 0. In columns 3 and 4 I look at whether the effect of studying triple science on the probability of choosing science at KS5 is larger for schools with a sixth form or not. The effect is identical. If part of the effect I find in my results is mechanical, then it would be stronger for schools also offering KS5 exams.

Another correlated concern is that taking triple science could potentially directly affect the possibility of being admitted to STEM majors at university and to science courses at A levels, 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 application choices. Still, while universities often require some A level subjects in order to admit students to certain majors, in no cases they require KS4 subjects. For instance, in 2013, an A level in math was required in 13% of the cases (i.e. of major-university combinations) and at least one A level in science was required in 12% of the cases. In no case in 2013 there was a specific requirement for age 14 (KS4) subjects.[36]

Finally, one may worry that the simple fact of having the possibility of being enrolled in advanced science but havng been excluded, for example because the class was oversubscribed and teachers had to select students, may generate a direct effect on some students. Figure 7 plots the distribution of the number of students in triple science classes in each school. From the Figure it is clear that schools bunch 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 teachers had to select students. Table 11 address this point by running my main specification (using as first stage equation 5) on the sample of schools where the triple science course was most likely not to be oversubscribed.Those schools where the number of students enrolled in the triple science classes was between 28 and 32 or between 58 ans 62.

4.5 Second Instrument

Table 12 shows results obtained from my second identification strategy, described by Equation 6.[37] The first columns does not include neighbourhood fixed effects, but controls for the

[35] If a school provides both KS4 and KS5 courses, it is said it has a "sixth form".

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

[37] 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 do not have information on whether they graduated only for the year 2007, so I only analyze effects on enrollment and on A level outcomes.

lagged value of my instrument: it compares neighbouroods which, the previous year had the same share of reachable schools offering the triple science course. The second column includes neighbourhood fixed effects.

This instrument compares students attending different schools which offer or do not offer triple science. However, the probability of 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. I address this point in Column 3, where I control whether my results are robust to the inclusion of the average quality of the set of reachable schools in year t as a control. I proxy school quality using the average school value added in the previous years. The results confirm the robustness of the first strategy: the estimated effects are very similar to those found in Table 3.

5 Mechanisms

This Section digs into the mechanisms that may generate the effect found in Section 4 and explores whether the effects obtained are actually generated by changes in curriculum or, since the treatment has multiple components, they are actually driven by changes in the peer composition of the courses attended or in the type of teachers in the school.

5.1 Peers

First, I analyze the peers channel. In particular, I use the following measure of peer quality in science (Q_{ist}) for student i in school s and year t who takes science courses D_{ist}:[38]

$$Q_{ist} = \overline{X}^{D}_{(-i)st} \tag{9}$$

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

The first panel of Figure 8 shows how peers' composition in the science course taken at age 14 changes after the introduction of the triple science option. 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 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 the courses taken, I control for peer quality in equation 4. 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 changes in peers' composition within-school over-time (following Hoxby [2000]). In particular, I use the fact that classes in the UK cannot be larger than 30.[39] Figure 7 plots the distribution of the number of students in triple science classes in each school. From the Figure it is clear that schools bunch at multiples of 30. There is a discontinuity both corresponding to 30 students and corresponding

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

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

to 60 students. I therefore predict, based on predetermined characteristics,[40] 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 will be as follows:

$$Q_{ist} = \theta_1 z_{st} + \theta_2 \widehat{Q^{top30}_{st(-i)}} + \theta_3 \widehat{Q^{others}_{st(-i)}} + \theta_4 \widehat{Q^{top30}_{st(-i)}} * z_{st} + \theta_5 \widehat{Q^{others}_{st(-i)}} * z_{st} + \theta_5 X_{ist} + \theta_s + \theta_t + \eta_{istj} \quad (10)$$

where $\widehat{Q^{top30}_{st(-i)}}$ 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 33.

5.2 Teachers

I CAN'T INCLUDE THE TABLE NOW BUT THE COEFFICIENTS WERE NOT SIGNIFICANT

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

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 in high school has no clear effect on high school graduation and university enrollment on average, but it has a positive effect for low income/high ability students. Moreover it shifts very high ability students towards high quality ivy-league universities. In particular stronger science preparation increases by 12% the likelihood of attending a ivy-league university in the UK for students in the top 20 percentiles of the distribution of grades in primary school.

More science courses at age 14 significantly increase the probability of enrolling and, very importantly, of graduating into a STEM major and scientific majors at university. This effect masks substantial gender heterogeneity: both boys and girls opt to traditionally higher quality courses on average, but girls choose more female-dominated subjects like medicine, instead of engineering and math.

[40]KS2 and KS3 science grades (both teacher assessed and from standardized exams) , gender, FSME.

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 and that differences in high schools and in the curriculum they offer explain some of the large differences in major choices, educational outcomes, and earnings, by SES status.

References

Joseph G. Altonji. The effects of high school curriculum on education and labor market outcomes. *The Journal of Human Resources*, 30(3):pp. 409–438, 1995.

Joseph G. Altonji, Erica Blom, and Costas Meghir. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. Technical Report 17985, National Bureau of Economic Research, Inc, April 2012.

Massimo Anelli and Giovanni Peri. Peer gender composition and choice of college major. Working Paper 18744, National Bureau of Economic Research, January 2013.

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. Cep discussion papers, Centre for Economic Performance, LSE, June 2013.

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 Informatrion Technology and Innovation Foundation, October 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, January 2003.

Magali Beffy, Denis Fougere, and Arnaud Maurel. Choosing the field of study in postsecondary education: do expected earnings matter? *Review of Economics and Statistics*, 94(1):334–347, 2012.

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

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.

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.

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, August 2012.

European Commission. The european commission: Europe 2020 strategy. Technical report, 2010.

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, April 2015.

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. 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.

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, July 2013.

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

Information Technology and Innovation Foundation. Refueling the u.s. innovation economy: Fresh approaches to stem education. Technical report, 2010.

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.

Ning Jia. Do stricter high school math requirements raise college stem attainment? Working Papers 8983, Mimeo, 2014.

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

Juanna Schrøter Joensen and Helena Skyt Nielsen. Mathematics and gender: Heterogeneity in causes and consequences. *The Economic Journal*, pages n/a–n/a, 2015.

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

Cortes Kalena, 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.

Lars J. Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of study, earnings and self-selection. Discussion Papers 794, Research Department of Statistics Norway, January 2015.

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, October 2010.

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, April 1995.

Wiswall M and B Zafar. Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, 2015.

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.

Ronni Pavan and Josh Kinsler. The Specificity of General Human Capital: Evidence from College Major Choice. *Journal of Labour Economics*, 2015.

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

Andrew Rendall and Michelle Rendall. Math matters: education choices and wage inequality. ECON - Working Papers 160, Department of Economics - University of Zurich, May 2014.

Marianne E. Page Scott E. Carrell and James E. West. Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144, 2010.

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 (IZA), October 2014.

Figures

Figure 1: Take up in triple science

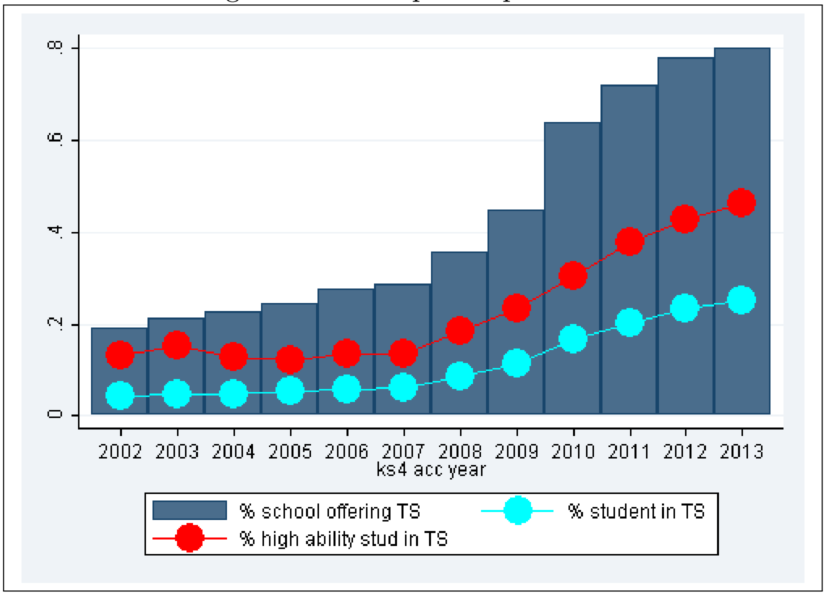


Figure 2: English Educational system, timeline

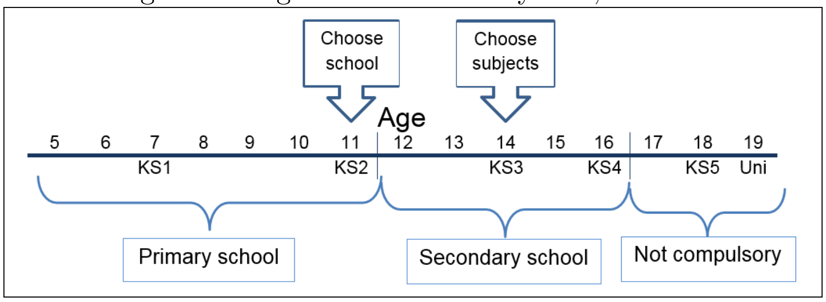
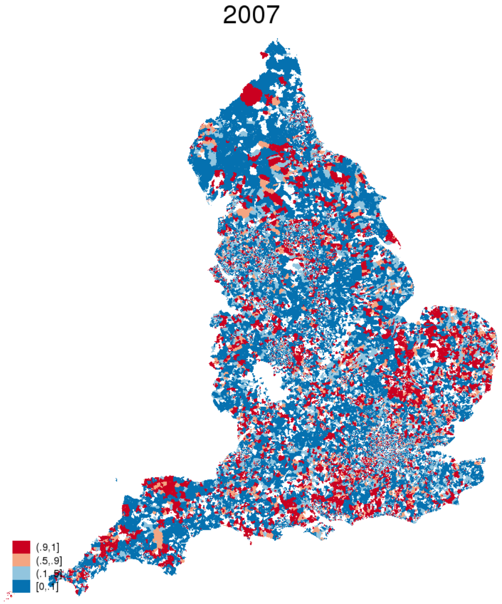


Figure 3: growth, spatial

offering in 2005 (exam taken in 2007)



offering in 2008

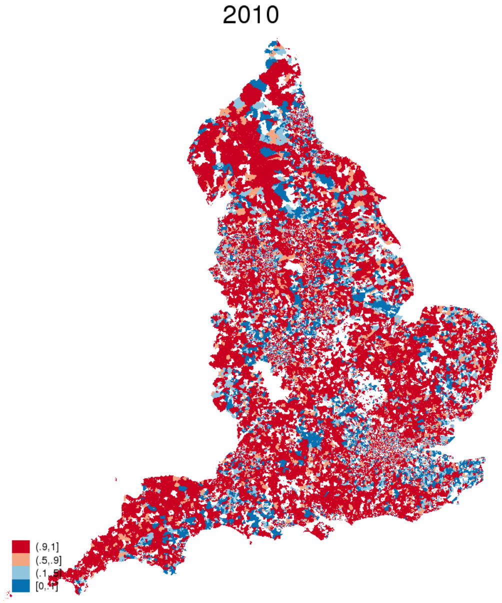


Figure 4: Instrument 2

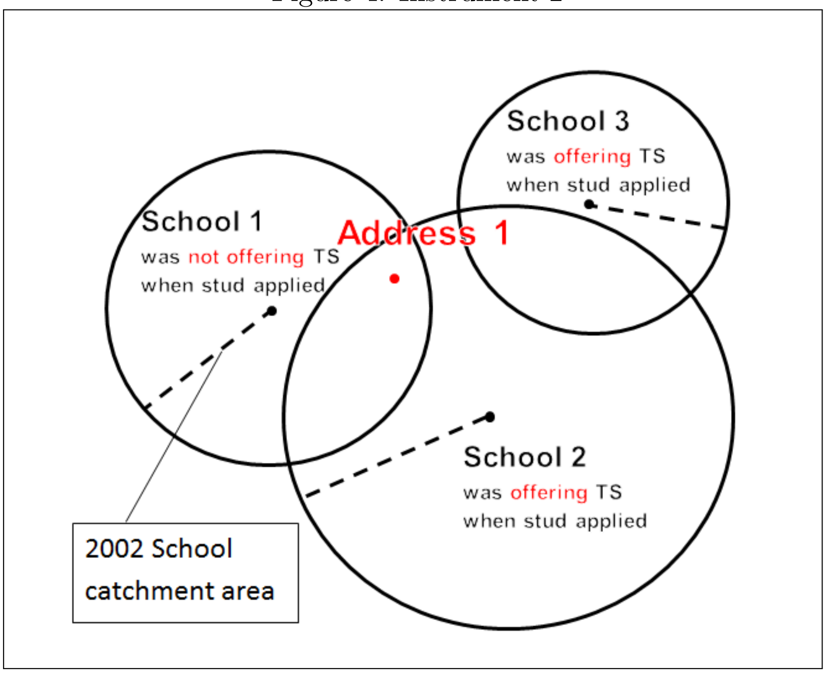
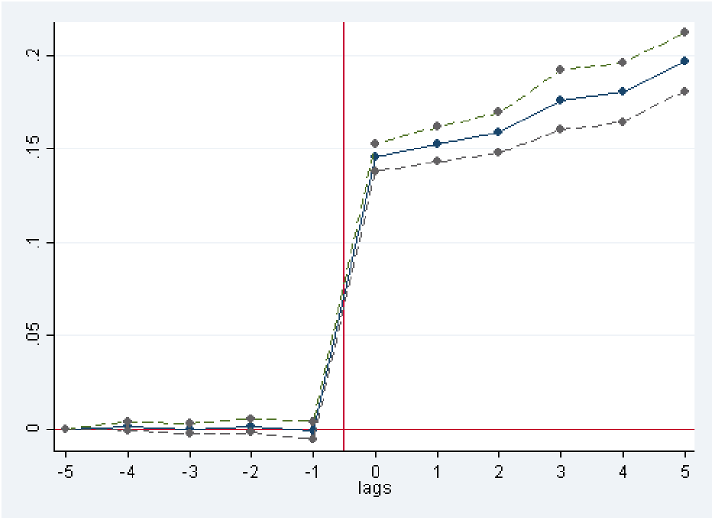
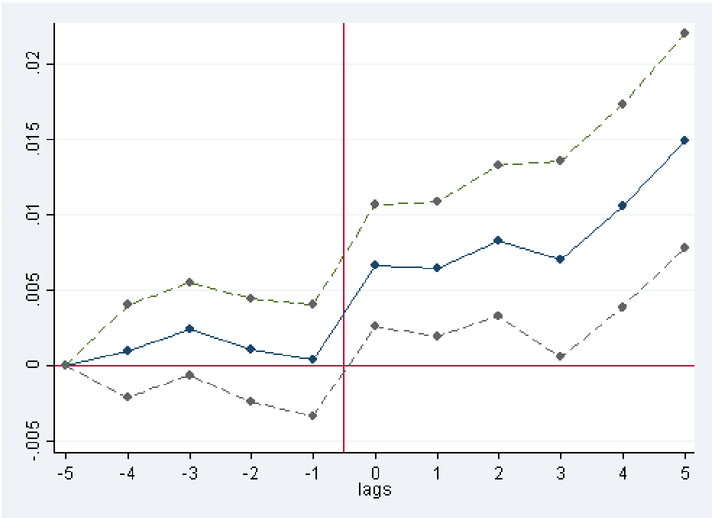


Figure 5: Parallel Trends: Leads and Lags of the instrument

Dep var: 1=Triple Science



Dep var: 1=A lev Science



Dep var: 1=Uni STEM

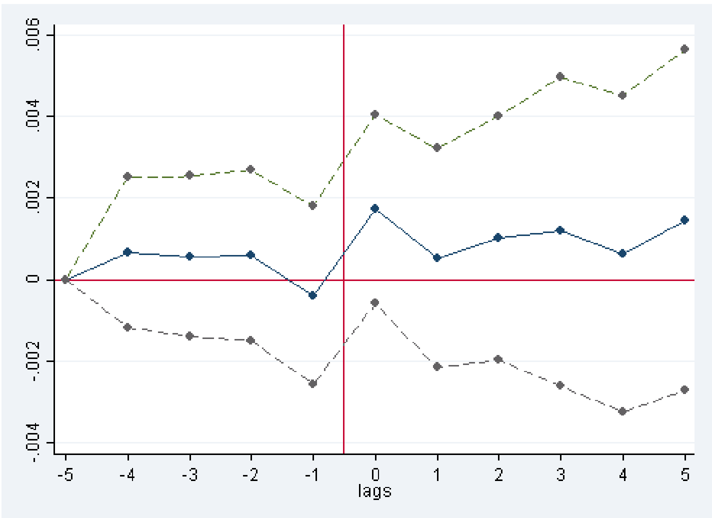
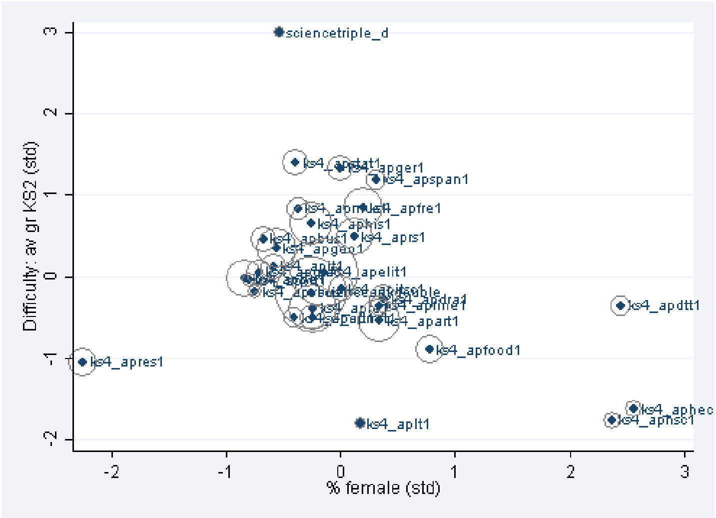
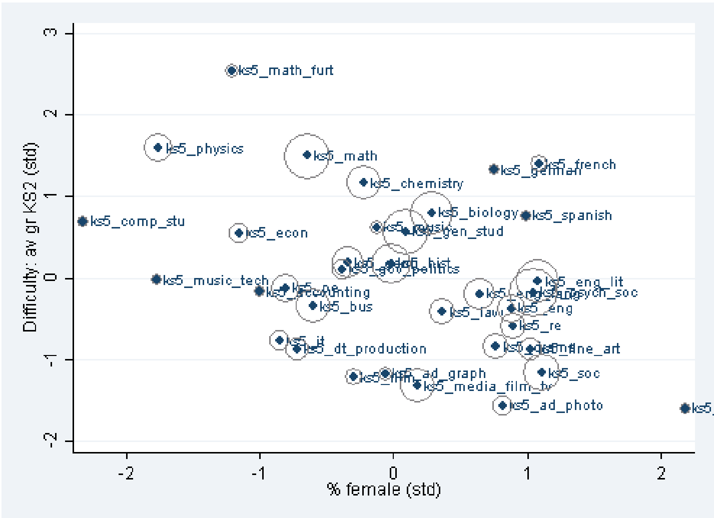


Figure 6: Subject descriptives

KS4 courses (age 14)



KS5 courses (age 16)



University courses (age 18)

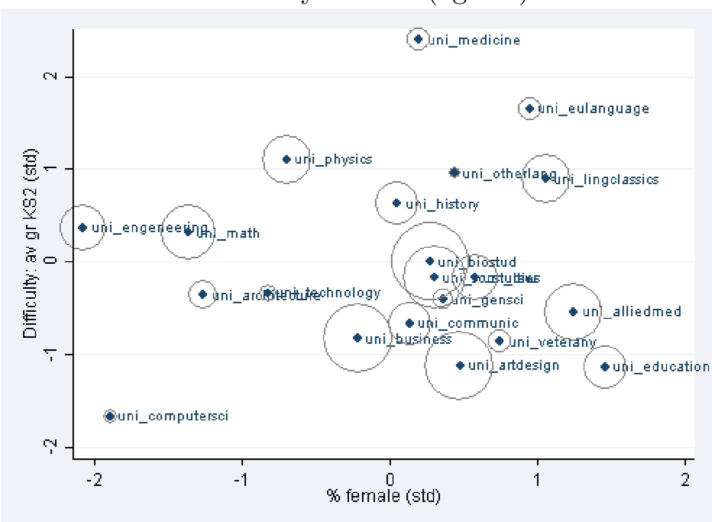


Figure 7: Class size and number of students in triple science

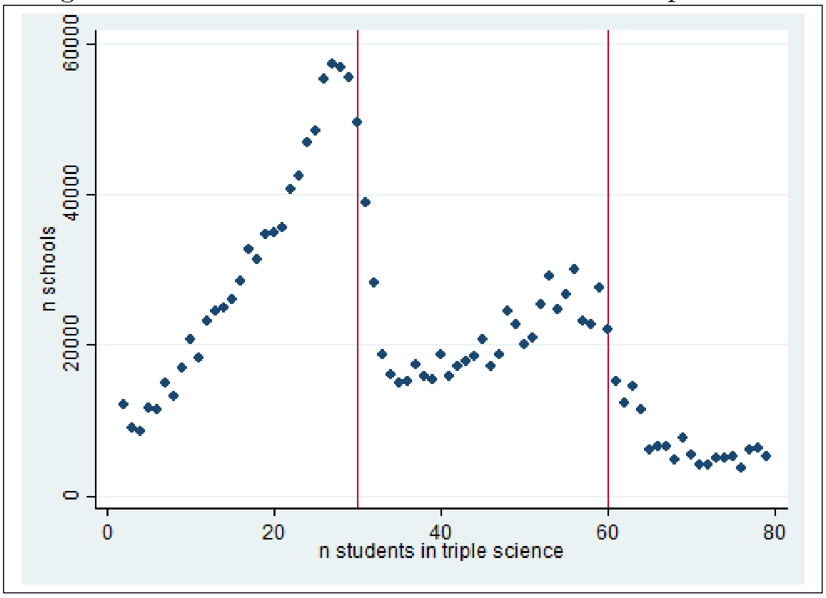
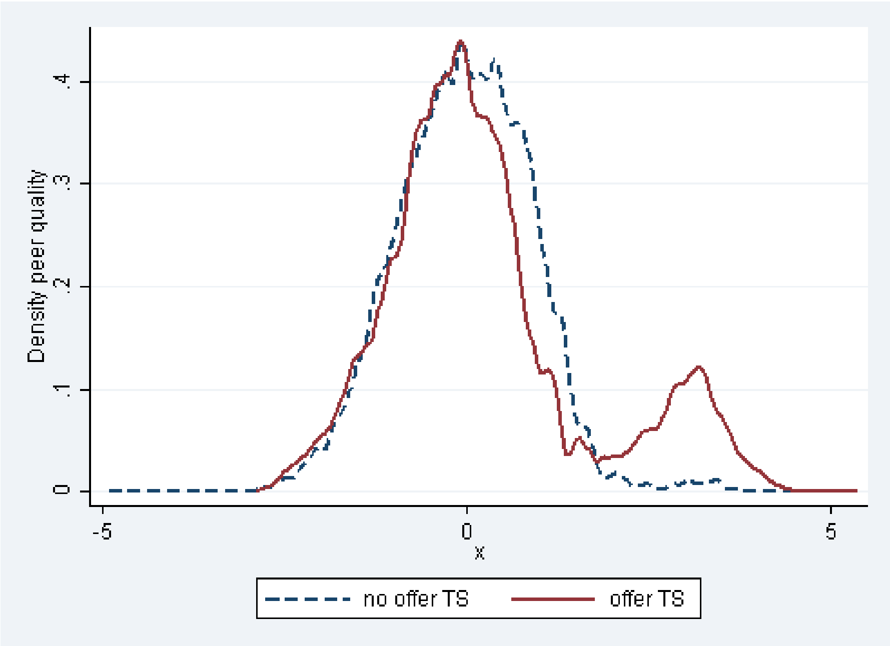
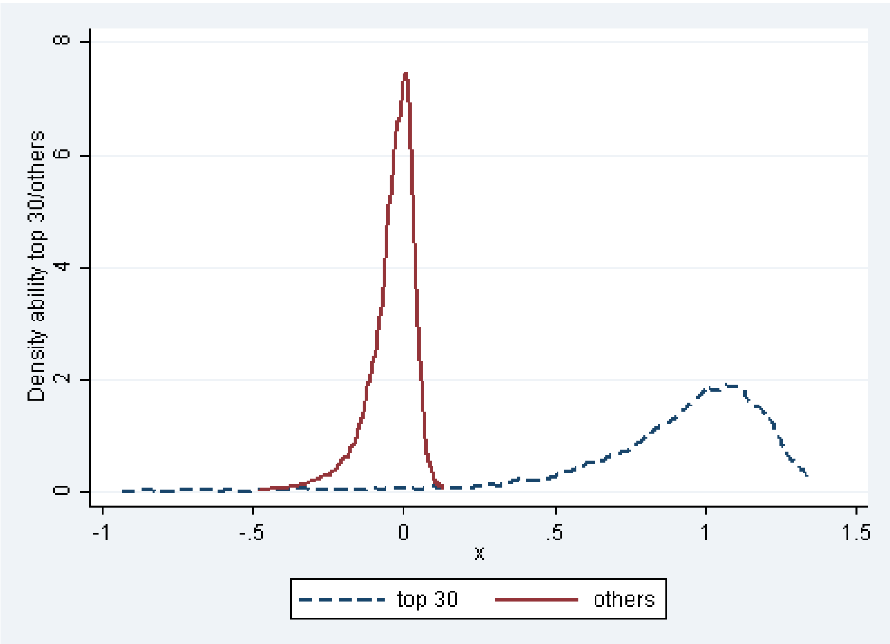


Figure 8: Peers

Actual peers' quality



Instrument



Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.	N
	Key Stage 4		
offer TS (unexpected)	0.196	0.397	2888369
1=Triple Sci	0.076	0.264	2994149
1=Double Sci	0.764	0.425	2994149
1=Single Sci	0.163	0.369	2994149
	Key Stage 5		
1=A lev sci (if KS5)	0.198	0.282	2994149
1=A lev math (if KS5)	0.142	0.252	2994149
	University		
1=uni	0.348	0.470	2505008
1=STEM (if uni)	0.126	0.198	2994149
1=Russell	0.046	0.211	2994149
1=graduate (in uni)	0.481	0.361	2005829
	Demographics		
1=female	0.497	0.500	2994149
1=FSM eligible	0.144	0.356	2963763

Robust standard errors clustered by school in parentheses.
controls years dummies school fixed effect.

Table 2: Main Results

	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)	
prim 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)	
prim 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

Robust standard errors clustered by school in parentheses. Additional controls years dummies. Student controls: gender, SES, Special Education Needs, primary school grade in science, math and english. Schools controls share of girls, share of low SES students, average primary school grade.

Table 3: Other outcomes

	Main Results					
	[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	-0.097**	0.438**	-0.021	
	(0.031)	(0.028)	(0.041)	(0.210)	(0.022)	
N	1332413	1339792	1690325	1690451	860615	
ymean	0.022	0.021	4.397	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***	0.009
	(0.010)	(0.005)	(0.004)	(0.003)	(0.005)	(0.007)
N	1690451	1690451	1690451	1690451	1690451	1690451
ymean	0.509	0.056	0.040	0.026	0.065	0.107
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)	(0.008)	(0.011)	
N	966777	966777	966777	966777	966777	
ymean	0.318	0.207	0.046	0.019	0.034	

Robust standard errors clustered by school in parentheses. Additional controls: School fixed effects, years dummies and usual students controls.

[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 2004-2008 sample because otherwise there will be no information on the graduation outcomes.

Table 4: Effects on other subjects

Dep var:	Difficulty (av KS2 grade)	% female
Δ ks5 courses	0.174***	-0.039**
	(0.020)	(0.019)
Δ ks5 sci courses	0.110***	-0.039***
	(0.009)	(0.007)
Δ ks5 no sci courses	0.065***	-0.000
	(0.018)	(0.018)
Δ uni majors	0.021***	-0.009
	(0.008)	(0.009)

Each line represents a different regression. Usual controls. Robust standard errors clustered at the school level. Additional controls: years dummies and usual students controls, school fixed effects.

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 sciemce grade prim schol			
		subgroup: 1st quint	
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 quint	
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 quint	
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 quint	
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 quint	
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:1=FSM	
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

Robust standard errors clustered by school in parentheses.
Additional controls: years dummies and usual students controls.

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 grade science prim 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)	(0.01)	(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

Robust standard errors clustered by school in parentheses. Additional controls: years dummies and usual students controls.

Table 7: Effects on other subjects by gender

Dep var:	% female		Difficulty (av KS2 grade)	
	Girls	Boys	Girls	Boys
Δ ks5 courses	0.125***	0.213***	-0.022	-0.051***
	(0.029)	(0.024)	(0.029)	(0.021)
Δ ks5 sci courses	0.083***	0.131***	-0.017***	-0.057***
	(0.011)	(0.012)	(0.006)	(0.010)
Δ ks5 no sci courses	0.042	0.082***	-0.005	0.006
	(0.027)	(0.020)	(0.028)	(0.018)
Δ uni majors	0.018*	0.020**	0.004	-0.019
	(0.011)	(0.009)	(0.012)	(0.012)

Each line represents a different regression. Usual controls. Robust standard errors clustered at the school level. Additional controls: years dummies and usual students controls, school fixed effects.

Table 8: Selection

	av KS2 gr	sci KS2 gr	1=FSM	1=A lev 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

Robust standard errors clustered by school in parentheses. controls years dummies school fixed effect.

Table 9: Balancing Tests

	OLS	OLS-Fe	Altonji	IV	IV-Fe	IV-Fe tr
	[1]	[2]	[3]	[4]	[5]	[6]
Dep var:			1=Average Grade prim school			
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

Robust standard errors clustered by school in parentheses. controls years dummies.

Table 10: Robustness: offer KS5 Science

	Sch level regr (offer)		Stud in schools wo sixth form	
Dep var:	1=Offer A lev	1=offer A lev	All schools	no 6 form
	Science	Math	Dep var: 1=A lev Sci	
	[1]	[2]	[3]	[4]
offertriple0	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

Robust standard errors clustered by school in parentheses. Additional controls: years dummies and usual students controls, school fixed effects.

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
ymean						

Robust standard errors clustered by school in parentheses. Additional controls: years dummies and usual students controls, school fixed effects.

Table 12: Identification based on share of reachable schools offering TS

	IV [1]	IV Neighb FE [2]	IV Neighb FE [3]
A lev Physics			
1=TP	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)
N	2860812	2861393	2861393
A lev Chemistry			
1=TP	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)
N	2860812	2861393	2861393
A lev Biology			
1=TP	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)
N	2860812	2861393	2861393
Uni Engineering			
1=TP	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)
N	2860812	2861393	2861393
Uni Medicine			
1=TP	-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

- Robust standard errors clustered by neighbourhood in parentheses. Additional controls: years dummies and usual students controls and some neighbourhood controls (av grade in primary school in science, math, english, share of girls, share of low SES).

Table 13: Peers

Dep var:	q peer sci[a]	1=KS5 sci	1=Russell	1=STEM	1=medicine	1=grad	1=grad STEM
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Z offer*ks2 sci q1	-0.095***						
	(0.011)						
Z offer*ks2 sci q2	-0.060***						
	(0.008)						
Z offer*ks2 sci q3	-0.031***						
	(0.007)						
Z offer*ks2 sci q4	0.024***						
	(0.007)						
Z offer*ks2 sci q5	0.055***						
	(0.007)						
Z offer*ks2 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)
qual peer (std)		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

Robust standard errors clustered by school in parentheses. Additional controls: years dummies and usual students controls, school fixed effects. F statistic: 24 (for column 2 and 3), 32 (for column 4-5-6).
[a] quality (based on science grade in ks2 (age 11) of peers in the same science class.

Appendices

Table 14: Balancing Tests

	RF	RF	IV	IV
	[1]	[2]	[3]	[4]
Dep var:	1=Grade Science prim school			
Z_{st}	0.006	0.011**		
	(0.004)	(0.005)		
1=TS			0.037	0.036
			(0.024)	(0.024)
N	1690451	1690451	1690451	1690451
ymean	0.015	0.015	0.015	0.015
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

Robust standard errors clustered by school in parentheses.
controls years dummies.

Table 15: Effect on other KS4 Subjects

Dep. var	All		Girls		Boys	
	Coeff.	Se	Coeff.	Se	Coeff.	Se
1=TS	1.000***	(0.000)	1.000***	(0.000)	1.000***	(0.000)
gcse (dummy) - eng lit	0.068**	(0.030)	0.075**	(0.030)	0.061*	(0.032)
gcse DT food	-0.027*	(0.016)	-0.047**	(0.024)	-0.009	(0.013)
gcse DT graphics	-0.015	(0.014)	-0.002	(0.017)	-0.027	(0.017)
gcse DT material	-0.014	(0.014)	0.000	(0.011)	-0.024	(0.022)
gcse DT textile	-0.006	(0.010)	-0.014	(0.020)	0.003	(0.002)
gcse (dummy) - art design	-0.008	(0.019)	0.001	(0.025)	-0.015	(0.019)
gcse (dummy) - history	-0.032*	(0.019)	-0.045*	(0.023)	-0.022	(0.021)
gcse (dummy) - geogr	0.007	(0.020)	0.010	(0.024)	0.005	(0.022)
gcse (dummy) - french	-0.015	(0.028)	-0.010	(0.033)	-0.020	(0.027)
gcse (dummy) - german	-0.065***	(0.018)	-0.072***	(0.022)	-0.060***	(0.018)
gcse (dummy) - business	-0.012	(0.019)	-0.012	(0.020)	-0.014	(0.021)
gcse (dummy) - religion	0.038	(0.036)	0.038	(0.043)	0.038	(0.034)
gcse (dummy) - religion (short)	0.056	(0.053)	0.058	(0.059)	0.055	(0.053)
gcse (dummy) - physical edu	0.033	(0.021)	0.035	(0.023)	0.031	(0.025)
gcse (dummy) - drama	0.007	(0.014)	-0.001	(0.020)	0.013	(0.014)
gcse (dummy) - inf tech	-0.034	(0.031)	-0.020	(0.032)	-0.048	(0.035)
gcse (dummy) - inf tech(short)	-0.027	(0.049)	-0.041	(0.060)	-0.014	(0.045)
gcse (dummy) - spanish	0.012	(0.015)	0.013	(0.020)	0.011	(0.013)
gcse (dummy) - music	-0.001	(0.008)	-0.012	(0.011)	0.009	(0.010)
gcse (dummy) - statistics	0.011	(0.034)	0.010	(0.038)	0.011	(0.034)
gcse (dummy) - media	-0.012	(0.022)	-0.016	(0.025)	-0.009	(0.023)
gcse (dummy) - fine art	0.005	(0.014)	0.007	(0.019)	0.004	(0.013)
gcse (dummy) - office technology	0.016	(0.028)	0.008	(0.032)	0.022	(0.028)
gcse (dummy) - home economics	0.000	(0.010)	-0.001	(0.021)	0.002*	(0.001)
gcse (dummy) - Applied buss	-0.001	(0.014)	-0.004	(0.015)	0.000	(0.015)
gcse (dummy) - health care	0.003	(0.011)	0.009	(0.022)	-0.002	(0.004)
gcse (dummy) - leisure	0.002	(0.008)	-0.003	(0.011)	0.007	(0.008)
gcse (dummy) - applied IT	-0.009	(0.021)	-0.009	(0.021)	-0.008	(0.024)

Each line represents a different regression. Usual controls. Robust standard errors clustered at the school level.

Table 16: Effect on other KS5 Subjects

Dep. var	All		Girls		Boys	
	Coeff.	Se	Coeff.	Se	Coeff.	Se
ks5_biology	0.035***	(0.005)	0.037***	(0.008)	0.034***	(0.006)
ks5_chemistry	0.037***	(0.004)	0.032***	(0.006)	0.040***	(0.005)
ks5_physics	0.025***	(0.003)	0.012***	(0.003)	0.036***	(0.005)
ks5_math	0.024***	(0.005)	0.016**	(0.007)	0.031***	(0.007)
ks5_comp_stu	-0.001	(0.001)	-0.001	(0.001)	-0.000	(0.003)
ks5_bus	-0.004	(0.005)	-0.004	(0.006)	-0.003	(0.006)
ks5_ad_texti	-0.003*	(0.002)	-0.005	(0.003)	-0.001*	(0.000)
ks5_fine_art	-0.001	(0.003)	-0.000	(0.005)	-0.002	(0.003)
ks5_hist	0.005	(0.005)	0.004	(0.008)	0.005	(0.006)
ks5_econ	0.003	(0.003)	0.002	(0.003)	0.004	(0.005)
ks5_re	-0.000	(0.004)	-0.001	(0.006)	0.001	(0.003)
ks5_law	-0.007**	(0.003)	-0.007	(0.005)	-0.008**	(0.004)
ks5_gov_politics	0.001	(0.003)	-0.005	(0.004)	0.007**	(0.003)
ks5_psych_soc	-0.010*	(0.006)	-0.015	(0.011)	-0.006	(0.005)
ks5_soc	0.003	(0.005)	0.001	(0.009)	0.004	(0.004)
ks5_eng	-0.002	(0.004)	-0.006	(0.007)	0.002	(0.004)
ks5_eng_lang	0.009*	(0.004)	0.014*	(0.007)	0.004	(0.004)
ks5_eng_lit	0.004	(0.006)	0.006	(0.010)	0.002	(0.004)
ks5_drama	-0.002	(0.003)	-0.008	(0.005)	0.003	(0.003)
ks5_media_film_tv	-0.012***	(0.005)	-0.013*	(0.007)	-0.011**	(0.005)
ks5_film	-0.003	(0.002)	-0.001	(0.003)	-0.005*	(0.003)
ks5_french	0.001	(0.002)	0.001	(0.004)	0.001	(0.002)
ks5_german	-0.003**	(0.001)	-0.002	(0.002)	-0.003**	(0.001)
ks5_music_tech	-0.004***	(0.001)	-0.001	(0.001)	-0.008***	(0.002)
ks5_pe	-0.005	(0.003)	-0.007*	(0.004)	-0.004	(0.005)
ks5_accounting	-0.002*	(0.001)	-0.002	(0.002)	-0.002	(0.002)
ks5_dt_production	-0.009***	(0.003)	-0.012***	(0.004)	-0.007*	(0.004)

Each line represents a different regression. Usual controls. Robust standard errors clustered at the school level.

Table 17: Effect on other university majors

Dep. variables	All		Girls		Boys	
	Coeff.	Se	Coeff.	Se	Coeff.	Se
uni_architecture	-0.003***	(0.001)	-0.002*	(0.001)	-0.004**	(0.002)
uni_medicine	0.002	(0.001)	0.003	(0.002)	0.001	(0.001)
uni_alliedmed	0.004*	(0.002)	0.008*	(0.004)	0.000	(0.002)
uni_biostud	-0.001	(0.003)	-0.001	(0.005)	-0.002	(0.004)
uni_vetagri	-0.001	(0.001)	-0.001	(0.002)	0.000	(0.001)
uni_physics	0.006***	(0.002)	0.001	(0.003)	0.009***	(0.003)
uni_math	0.001	(0.002)	-0.002	(0.002)	0.003	(0.004)
uni_engineering	0.007***	(0.002)	0.003**	(0.001)	0.011***	(0.003)
uni_computersci	-0.001	(0.001)	-0.001	(0.001)	-0.000	(0.002)
uni_techn	-0.000	(0.001)	-0.000	(0.001)	-0.000	(0.001)
uni_socstudies	0.003	(0.003)	0.005	(0.005)	0.001	(0.003)
uni_law	-0.004*	(0.002)	-0.006*	(0.003)	-0.002	(0.002)
uni_business	0.001	(0.003)	0.001	(0.004)	-0.000	(0.004)
uni_communic	0.000	(0.002)	0.001	(0.003)	-0.001	(0.002)
uni_lingclassic	0.005**	(0.002)	0.004	(0.004)	0.006***	(0.002)
uni_eulang	-0.000	(0.001)	-0.000	(0.002)	-0.000	(0.001)
uni_otherlan	0.000	(0.000)	-0.000	(0.001)	0.000	(0.001)
uni_history	0.001	(0.002)	0.003	(0.003)	-0.001	(0.002)
uni_artdesign	-0.000	(0.003)	0.001	(0.005)	-0.002	(0.003)
uni_education	-0.001	(0.002)	-0.001	(0.004)	-0.001	(0.001)
uni_gensci	-0.000	(0.001)	-0.001	(0.002)	0.000	(0.001)

Each line represents a different regression. Usual controls. Robust standard errors clustered at the school level.