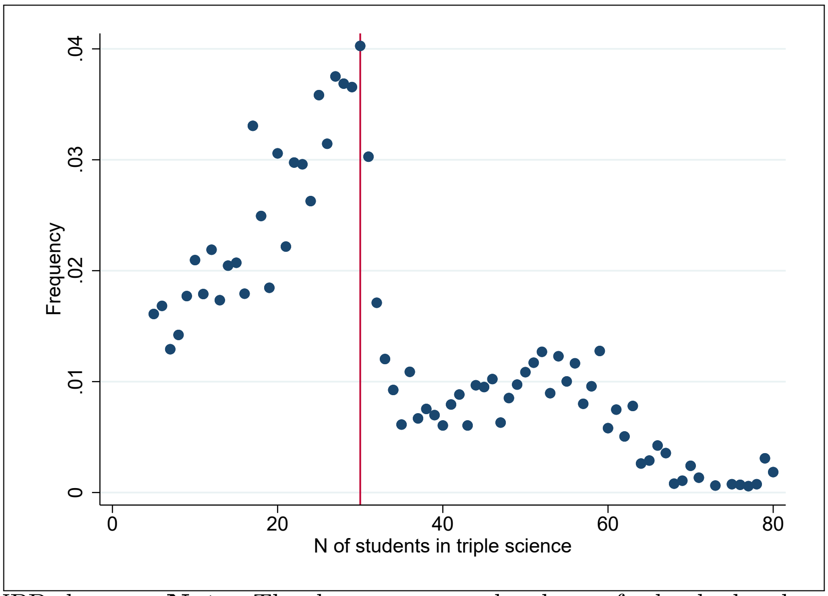
STEM Graduates and Secondary School Curriculum: Does Early Exposure to Science Matter?

Marta De Philippis*

Online Appendix

A Additional tables and figures

Figure A1: Class size and number of students in triple science



Source: NPD dataset. **Note**: The dots represent the share of schools, by the size of the triple science class. Since the maximum class size in England is usually 30, this Figure shows that there is a certain degree of rationing of students in triple science classes (in schools where class size in TS is close to 30), but most schools do not experience rationing.

*Bank of Italy and Centre of Economic Performance; email: marta.dephilippis@bancaditalia.it

Table A1: Robustness: Exclusion restriction

	[1]	[2]	[3]	[4]	[5]	[6]
Panel 1: KS5 (age 17) outcomes						
Dep var:	KS5	Biology	Chemistry	Physics	Math	N KS5 sci
TS	-0.012	0.041***	0.039***	0.021***	0.025***	0.101***
	(0.010)	(0.005)	(0.004)	(0.003)	(0.006)	(0.010)
N	1,652,391	1,652,391	1,652,391	1,652,391	1,652,391	1,652,391
Panel 2: University outcomes, enrollment						
Dep var:	uni	Russell	STEM	Engin.	Physics	Medicine
TS	0.021**	0.012***	0.013**	0.007***	0.006**	0.006**
	(0.009)	(0.004)	(0.006)	(0.002)	(0.002)	(0.003)
N	1,652,391	1,652,391	1,652,391	1,652,391	1,652,391	1,652,391
Panel 3: University outcomes, graduation						
Dep var:	grad	Russell	STEM	Engin.	Physics	Medicine
TS	0.056**	0.020	0.029*	0.009	0.018***	0.010
	(0.028)	(0.012)	(0.016)	(0.006)	(0.007)	(0.008)
N	955,160	955,160	955,160	948,328	948,328	955,160

Note: Medicine includes medicine and subjects allied to medicine; STEM refers to Physical science, Mathematical and Computer science, Medicine and Engineering. The dummies on university enrollment and graduation refer to students who enroll on time (no years off) and graduate on time. Russell indicates whether students enroll or graduate in a university belonging to the Russell group. All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. The sample consists of students enrolled in maintained secondary schools in England, taking their final KS4 exams (at age 16) between 2005 and 2010 (panels 1 and 2) and between 2005 and 2008 (panel 3), who apply at age 11 to schools not offering triple science. Only for schools not rationing students in TS classes, that is where the number of TS students is not around 30 (< 28 and > 32). Additional controls: year and school fixed effects; student controls: gender, Free School Meal Eligible, Special Education Needs, primary school grade in science, math and English; schools controls: school size. Robust standard errors, clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table A2: Offering of KS5 science courses

	[1]	[2]	[3]	[4]	[5]
Dep var:	School level probability of offering KS5 courses in:				
	$\geq$ 1 Science	Chemistry	Biology	Physics	Math
Offer TS	0.002	0.001	0.002	0.004	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	10,855	10,855	10,855	10,855	10,855

Note: The table looks at whether schools that start to offer triple science at KS4 are more likely to also start to offer new science courses at KS5. All dependent variables are set equal to 0 if schools do not offer KS5 courses or do not offer KS5 courses in science or math. The sample consists of the maintained secondary schools in England attended by the students in my regression sample. Additional controls: year and school fixed effects, school size, share of female, average KS2 grade in science, math and english, control for whether the school also offers KS5 courses or not. Robust standard errors, clustered by school in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table A3: Effect of taking TS on other KS5 and university subjects

Dep. var	All		Girls		Boys	
	Coeff.	se	Coeff.	se	Coeff.	se
			KS5 Subjects (age 17)			
Business	-0.003	(0.005)	-0.005	(0.006)	-0.002	(0.006)
History	0.004	(0.005)	0.004	(0.008)	0.004	(0.006)
Law	-0.006*	(0.003)	-0.007	(0.005)	-0.006*	(0.003)
English	0.009**	(0.005)	0.015**	(0.007)	0.005	(0.004)
Drama	-0.003	(0.003)	-0.010*	(0.005)	0.003	(0.003)
Media	-0.015***	(0.005)	-0.015**	(0.007)	-0.015***	(0.005)
Movies	-0.004*	(0.002)	-0.000	(0.003)	-0.007**	(0.003)
German	-0.003**	(0.001)	-0.003	(0.003)	-0.003**	(0.001)
Spanish	0.002	(0.002)	0.003	(0.003)	0.002	(0.001)
Music	-0.003*	(0.002)	-0.005**	(0.002)	-0.001	(0.002)
Music tech	-0.004***	(0.001)	-0.001	(0.001)	-0.006**	(0.002)
Physical Edu	-0.006*	(0.003)	-0.008**	(0.004)	-0.004	(0.005)
Accounting	-0.002*	(0.001)	-0.003*	(0.002)	-0.002	(0.002)
Product design	-0.010***	(0.003)	-0.013***	(0.004)	-0.008**	(0.004)
			University degrees (age 19)			
Architecture	-0.003**	(0.001)	-0.002	(0.001)	-0.004**	(0.002)
Arts and Design	-0.001	(0.003)	-0.003	(0.005)	0.000	(0.003)
Business	0.002	(0.003)	0.004	(0.004)	-0.001	(0.004)
Education	-0.001	(0.002)	-0.001	(0.004)	-0.002	(0.001)
European lit.	-0.001	(0.001)	-0.001	(0.002)	-0.000	(0.001)
History	0.001	(0.002)	0.004	(0.003)	-0.002	(0.002)
Linguistics	0.005**	(0.002)	0.004	(0.004)	0.006***	(0.002)
Mass communication	-0.000	(0.002)	0.001	(0.003)	-0.001	(0.002)
Other languages	-0.000	(0.000)	-0.001	(0.001)	0.000	(0.001)
Social studies	0.004	(0.003)	0.006	(0.005)	0.003	(0.003)

Note: Each line represents a different regression. The Table reports the effect on a selection of subjects only. Columns 1, 3 and 5 display the coefficients on the independent variable TS. All dependent variables are set equal to 0 if students do not continue studying or if they do not take that subject. The dummies on university enrollment refer to students who enroll on time (no years off). Usual controls. Robust standard errors clustered at the school level. N of observations: 1,696,336. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

B An alternative instrument

This Section presents an alternative identification strategy that addresses the concern that, even if there is evidence that schools do not decide when to offer triple science depending on observable characteristics of the cohorts currently in the school, unobservable characteristics may still matter. While this is impossible to test, I show that my findings are robust (even if often do not reach a conventional level of statistical significance) to an alternative instrument, which exploits variation in available courses that existed before the current students started to attend their secondary schools. This excludes the possibility that the choice of offering triple science depends on (unobservable) characteristics of the particular cohort of students 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 when schools in England are over-subscribed, they usually prioritize students based on geographical distance.[1] Therefore, in each year there will be a maximum home-to-school distance, above which students will not be accepted. I build my instrument in two steps: first, I compute the school catchment areas for each year as the area delimited by the circle whose centre is the school and ray is the maximum observed home-to-school distance,[2] 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 living at age 9 in neighbourhood p and belonging to cohort t applied to secondary school. Student address refers to the lower level output area (LLOA)[3] where student i used to live at age 9. 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 within each catchment area.

I estimate the following first stage equation:

$$TS_{ipt} = \theta_1 Z^2_{pt} + \theta_3 X_{ipt} + \theta_p + \theta_t + v_{ipt} \tag{1}$$

where TS_{ipt} is the usual dummy indicating whether student i in year t, who used to live in neighbourhood p when he/she was 9, takes triple science and 0 otherwise; Z^2_{pt} is the share of reachable schools offering triple science when students of cohort t, residing in

[1] With some exceptions for students with siblings attending the same school or for students with special education needs. Since I do not have the full set of information necessary to simulate the exact admission formula for each school, I cannot adopt a Regression Discontinuity Design.

[2] In order to exclude exceptions I eliminated some outliers (the distances higher than the 1th percentile for every school). My results are however robust to different definitions of outliers.

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

neighbourhood p at age 9, applied to secondary school; X_{ipt} are individual and school controls (gender, Free School Meal Eligibility, grades at KS2, controls for the composition of students in the school) and θ_t and θ_p are cohort and neighbourhood fixed effects respectively; v_{ipt} is the error term.

This instrument compares students attending schools that offer triple science with students attending schools not offering it, i.e. it uses across school-within neighbourhood-overtime variation (instead of within school-overtime variation). Offering triple science is however likely to be related to other school characteristics, like school quality, that may directly affect the choice of university degree. This issue is more relevant when using across school rather than within school variation because differences in quality across schools are more sizeable than differences within schools over time. I address this concern by including as control some measure of school quality (the average KS2 grade and the share of free school meal eligible among students attending it).

Table B1 shows the results obtained from my second identification strategy. Odd columns do not include neighbourhood fixed effects, but control for the lagged value of my instrument: they compare neighbourhoods which had the same share of reachable schools offering the triple science course the previous year and they exploit variation between t and $t-1$. All other columns include neighbourhood fixed effects.

The results confirm the robustness of the first identification strategy. The points estimates obtained through this strategy however often do not reach a conventional level of statistical significance, probably because of the smaller amount of variation in this second instrument, and are slightly larger, this may be related to the different type of exploited variation. Compliers for the first instrument are all individuals who take triple science because their school unexpectedly starts to offer it, this also includes very good students who happened to be enrolled in a school not offering triple science; compliers in the second instruments are instead students who take triple science thanks to a larger supply of the triple science course in the set of reachable schools in their neighbourhood. In this second case, very good students would probably have enrolled in a school offering triple science in any case. This suggests compliers for the second strategy exclude the extremely high ability students.

Table B1: Identification based on changes in catchment areas

Dep. Var.:	KS5 Science		STEM enroll		STEM grad	
	[1]	[2]	[3]	[4]	[5]	[6]
TS	0.122***	0.070	0.013	0.049	0.020	0.082
	(0.019)	(0.078)	(0.012)	(0.055)	(0.014)	(0.132)
Z^2_{pt-1}	0.005*		-0.002		-0.005**	
	(0.003)		(0.002)		(0.002)	
N	2,859,695	2,861,059	2,859,695	2,861,059	1,907,061	1,908,080
Neigh Fe	No	Yes	No	Yes	No	Yes
Controls sch. qual	No	Yes	No	Yes	No	Yes

Note: All dependent variables are set equal to 0 if students do not continue studying or if they do not take the considered subjects. The dummies on university enrollment and graduation refer to students who enroll on time (no years off) and graduate on time. STEM refers to Physics, Chemistry, Biology, Mathematical and Computer science and Engineering. All estimates use as instrument for TS Z^2_{pt}, the share of reachable schools for cohort t living in neighbourhood p offering TS. Z^2_{pt-1} refers to the share of reachable school in neighbourhood p that offer TS the year before cohort t applies. The sample consists of students enrolled in maintained secondary schools in England, taking their final KS4 exams (at age 16) between 2005 and 2010 (columns 1 to 4) and 2005 and 2008 (columns 5 and 6). Additional controls: year fixed effects; student controls: gender, Free School Meal Eligibility, primary school grade in science, math and English; school controls: school size, school quality (the average KS2 grade of students attending school s in cohort t). Robust standard errors, clustered by neighbourhood in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.