EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 25/03

February 2025

Math Exposure and University Performance: Causal Evidence from Twins

By

Graziella Bertocchi

(University of Modena and Reggio Emilia, EIEF, CEPR, CHILD, Dondena, GLO, and IZA)

Luca Bonacini

(University of Bologna and GLO)

Majlinda Joxhe

(University of Bologna and GLO)

Giuseppe Pignataro

(University of Bologna)

Math Exposure and University Performance: Causal Evidence from Twins[*]

Graziella Bertocchi[†]
Luca Bonacini[‡]
Majlinda Joxhe[§]
Giuseppe Pignataro[¶]

February 13, 2025

Abstract

We estimate the causal effect of exposure to math during high school on university major choice and performance, using a unique administrative dataset of 1,396 twins extracted from the entire student population enrolled between 2011 and 2021 at an Italian university. We apply a Twin Fixed Effect (TFE) estimator to account for unobserved factors like shared family background. We find that attending a low-math high school reduces the likelihood of enrolling in STEM majors by 32.6 percentage points and improves university performance, by increasing the likelihood of on-time graduation by 11.7 percentage points and boosting grades by 0.139 standard deviations. Leveraging a high school reform that expanded the math content in traditionally low-math curricula, we show that the added math background further reduces STEM enrollment for treated students, while it drives their improvement in performance. Our results suggest that, while increased math exposure does not necessarily boost STEM enrollment, it equips students with skills that help them improve their university outcomes. Compared with TFE, Ordinary Least Squares estimates of the effect of math exhibit a downward bias. The same applies to Difference-in-Differences estimates of the effect of the reform obtained using the entire student population.

JEL codes: D10; I21; I23; I28; J24.
Keywords: Math Exposure, Twins, Twin Fixed Effects, Major Choice, STEM, University Performance, High School Reform.

[*]Financial support: Almaidea Grant - Linea di intervento B. Project: Inequality of educational opportunities in Unibo. Special thanks to Patrizia Mondin and the ER.GO team – especially Annalisa Rambaldi and Valentina Sansivieri – for granting and processing the data, to the Datawarehouse Office at the University of Bologna—especially Camilla Valentini and Danilo Roberto Cinti—for accessing and organizing the administrative data, and to Cristina Specchi for outstanding research assistance.

[†]Department of Economics, University of Modena and Reggio Emilia, EIEF, CEPR, CHILD, Dondena, GLO, and IZA, Email: graziella.bertocchi@unimore.it.

[‡]Department of Economics, University of Bologna and GLO, Email: l.bonacini@unibo.it.

[§]Department of Economics, University of Bologna and GLO, Email: majlinda.joxhe@unibo.it

[¶]Department of Economics, University of Bologna, Email: giuseppe.pignataro@unibo.it

Significance

This study investigates the impact of exposure to math in high school on college major choice and academic performance, using a dataset of twins drawn from the student population of an Italian university. Our empirical approach accounts for confounding factors, such as family background, that would otherwise bias our results. We find that attending a low-math high school decreases the likelihood of choosing a STEM major, while it boosts academic performance in terms of on-time graduation and grades. A national high school reform that increased math hours in low-math curricula further reduced STEM enrollment but was the driver of the improvement in performance for treated students. Hence, more math does not increase STEM enrollment, but does provide students valuable skills.

1 Introduction

The relationship between math exposure in high school and academic and career outcomes has been widely studied, with suggestive evidence showing that advanced math coursework is associated with improvements in educational achievements, by increasing college enrollment, graduation rates, the choice of majoring in Science, Technology, Engineering, and Math (STEM), university performance, and subsequent earnings (Altonji, 1995; Long et al., 2009; Aughinbaugh, 2012). One issue with the available evidence, however, is its difficulty in establishing a causal relationship, because of the family background could confound the influence of math on subsequent outcomes.

Building on previous insights but introducing a different methodology, our study provides causal evidence on how high school math exposure shapes college major choice and academic performance. Thanks to a unique identifier, we extract a sample of 1,396 twins from the student population enrolled at the University of Bologna, Italy's second largest university, in the period 2011-2021. This dataset allows us to apply an empirical strategy based on a Twin Fixed Effect (TFE) estimator, to address unobserved confounders like family background that often bias Ordinary Least Squares (OLS) estimates (Behrman and Taubman, 1976; Griliches, 1979; Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998). By exploiting variation in high school attendance within twin pairs, this approach allows us to isolate the impact of prior math exposure on STEM major choice and two key academic outcomes, on-time graduation rates and grades. Our measure of math exposure is based on administrative information on the type of high school attended by each twin in the sample, where high schools are classified into two types, defined by the math content of their respective curricula.

In the first part of our investigation, we assess the effect of having attended a low-math school on students enrolled at university during the entire 2011-2021 period. We show that students from low-math schools are 32.6 percentage points less likely to pursue STEM majors, reinforcing previous research on the importance of rigorous math education for STEM pathways (Levine and Zimmerman, 1995; Rose and Betts, 2004; Joensen and Nielsen, 2009). However, in terms of performance, our results appear to contradict much of the existing literature (Ely and Hittle, 1990), as students with less intensive math exposure exhibit a higher likelihood of graduation on time, by 11.7 percentage points, and a 0.139 standard deviation boost in grades. Hence, we find no evidence that a stronger math background improves performance, as the literature suggests. These preliminary findings are corroborated by a number of robustness checks, comparisons, and extensions, including a heterogeneity analysis that accounts for gender, zygosity, ability, and major choice. The presence of a gendered dimension in major choice and the effects of math aligns with recent contributions by Joensen and Nielsen (2016), Delaney and Devereux (2019), Bertocchi et al. (2023), De Philippis (2023), Delaney and Devereux (2025).

In the second part of the paper, we examine the impact of a national high school reform, enacted in 2010 and affecting students enrolling at university in 2015, that increased math hours in low-math curricula, while still keeping them below those in high-math curricula. By comparing the TFE estimates obtained separately after restricting the sample to the pre- and post-reform periods, we obtain striking results. Regarding STEM major choice, we show that the reform led to a further 16.8 percentage point drop among treated students in the post-reform period. Hence, increased math exposure in low-math high schools made students even less interested in pursuing math-intensive university majors. Regarding the two indicators of performance, we find that their improvement is entirely driven by the reform itself, since no significant effects are evident before its implementation. This finding is confirmed even controlling for a composition effect due to the fact that treated students may opt for less demanding non- STEM majors. Hence, it is the stronger math background provided by the reform that drives the improvement in the performance of students attending low-math schools. Taken together, these results suggest that while enhanced math exposure does not make STEM majors more attractive for those students who had chosen a low-math track to begin with, it is effective in equipping them with skills that help them thrive in their university studies, regardless of their chosen field.

While twin studies have been employed to analyze multiple education-related issues (Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998; Behrman and Rosenzweig, 1999; Zhang et al., 2007; Lin and Liu, 2009; Rosenzweig and Zhang, 2009; Li et al., 2012; Rosenzweig and Zhang, 2013; Behrman et al., 2015), our study is unique in its use of twin data to investigate the drivers of major choice and academic performance.

Furthermore, a unique advantage of our specific twins sample is that it is drawn from the entire reference population of university students. Most twin analyses are performed instead on samples based on twin registries or local surveys, whose construction requires the voluntary participation of twins and an active recruitment effort by researchers, inducing a potential for volunteer or recruitment bias (Webbink et al., 2006). Using administrative education data to identify twins, our study demonstrates the potential of large population-based datasets for rigorous twin analyses, echoing similar methodologies applied in historical settings (Feigenbaum and Tan, 2020).

The occurrence of a reform during the sample period under examination presents an opportunity for an alternative identification strategy based on the Difference in Differences (DiD) method. However, the latter can only be applied to the entire student population, since it relies on pre- and post-reform variation that is absent for twins who graduate from high school in the same year. A comparison of our TFE estimates obtained from the twins sample with DiD estimates obtained from the entire student population underscores the superiority of the TFE approach, pointing to a downward bias in the DiD that is likely due to selection into school types.

2 Institutional Setting

The Italian educational system In Italy, students typically begin university in the year they turn 19, after eight years of primary school and five years of high school. While primary school curricula are standardized, high schools are divided into distinct tracks based on their focus on mathematics and science rather than humanities.[1] Importantly, university admission in Italy is not restricted by the high school track. Students from both high- and low-math backgrounds are free to pursue any field of study at the university level, allowing flexibility in academic and career choices regardless of their prior specialization. However, students are required to choose their major at the time of enrollment.

Performance assessment Performance assessment is based on exams. The grade is on a scale from 0 to 30, with 18 being the minimum passing score and 30 *cum laude* reserved for outstanding performance. Students are often allowed to reject grades and retake exams, whether or not they have achieved the passing score of 18, to improve their grade point average. Although degrees last for three years for undergraduate programs and five or six years for single-cycle programs, students are allowed to graduate later on, so that graduating on time can be taken as an additional measure of performance alongside grades.

The University of Bologna With a student body of approximately 80,000, the University of Bologna (Unibo) is Italy's second largest university and a cornerstone of the national public university system. Unibo provides a diverse range of degree programs in all fields of study. Its student population reflects a mix of backgrounds, drawing individuals from every corner of Italy: around 45% hail from the Emilia Romagna region (Unibo's home base), 25% from the south, 10% from the central regions, 10% from the north, and 10% are international students.[2]

The 2010 high school reform Starting with the 2010/2011 school year, a high school reform—known as the *Gelmini* Reform[3]—increased the number of hours devoted to mathematics in traditionally low-math schools, while the number of math hours did not change in traditionally high-math schools.[4] Students who enrolled at university in the academic year 2015/2016 represent the first cohort potentially affected by this reform.

[1]High-math tracks, such as the scientific *lyceum* and technical/vocational schools, emphasize mathematical, scientific, and technological disciplines. In contrast, low-math tracks prioritize the humanities and include the classical and linguistic *lyceum*, as well as teacher training and art schools. Low-math schools are centered on the humanities, while high-math schools focus more heavily on scientific and technical education.

[2]For a comparison of Unibo's students with the broader population of Italian university students, see Appendix A.

[3]For more information, see www.parlamento.it/leggi. Bertocchi et al. (2023) exploit the same reform to analyze how high school math affects the gender gap in Economics enrollment.

[4]The number of weekly math hours in first grade rose from two to three in the former type, while it remained at five in the latter. For more information, see https://leg16.camera.it.

3 Empirical Framework

3.1 Data

Our comprehensive longitudinal database provides detailed administrative records and individual characteristics for Italian-born students who graduated from high school the year they turned 19 between 2011 and 2021 and subsequently enrolled at Unibo.[5] Leveraging a unique identifier—the Italian *fiscal code*—we identified a sample of 1,396 twins from a population of 123,015 students, matching on family name, date of birth, and municipality of birth.[6] The dataset includes information on gender, high school tracks, high school exit grades, and municipality of residence. Our analysis focuses on three key outcomes: a student's likelihood of enrolling in a STEM degree program and two measures of academic performance, represented by on-time graduation[7] and grades (restricted to passing grades that have been accepted).[8] This rich dataset offers a unique opportunity to explore the relationship between high school experiences, educational choices, and university performance.

Table 1: Summary statistics - Twins and non-twins samples

	Twins			Non-twins		
	Obs.	Mean	S.D.	Obs.	Mean	S.D.
STEM	1396	0.430	0.495	121619	0.364	0.481
On-Time Graduation	1092	0.532	0.499	91545	0.516	0.500
Grades	20091	26.647	3.338	1638963	26.471	3.375
Low Math	1396	0.343	0.475	121619	0.371	0.483
Female	1396	0.590	0.492	121619	0.572	0.495
High School Exit Grade	1396	83.941	11.662	121510	83.287	11.780
High School Graduation Year	1396	2016.125	3.160	121619	2016.114	3.169
Area of Birth: E-R	1396	0.750	0.433	121619	0.548	0.498
Area of Birth: North	1396	0.060	0.238	121619	0.123	0.328
Area of Birth: Center	1396	0.067	0.251	121619	0.135	0.341
Area of Birth: South	1396	0.122	0.328	121619	0.194	0.395

Given the relatively small size of the twins sample and the fact that twins represent only 1.14 percent of the total Unibo student population, to assess the external validity of our findings it is worth considering how closely the twins sample reflects the broader student

[5]In more detail, by considering the students who graduated from high school the year they turn 19 we exclude those that repeated a year or more during high school. Although in the vast majority of cases students enroll at university right after high school graduation, four students in the sample (two of whom belong to a twin pair) graduated in 2021 but enrolled at Unibo in 2022. Excluding these four students from the sample has no impact on our results.

[6]The sample also includes ten groups of triplets, adding up to 30 observations. When we refer to twin pairs, we also include the 10 groups of triplets.

[7]For on-time graduation, the number of observations is restricted since we can only consider students enrolled by more than three years for undergraduate (three-year) degree programs and by more than five/six years for single-cycle (five/six-year) degree programs.

[8]The unit of observation is represented by all grades, rather than the average grade for each student, since collapsing the sample to average grades would prevent us from accounting for differences in the number of exams that students have passed.

body. Hence, in Table 1 we compare the twins sample of 1,396 twins with the sample of 121,619 non-twins enrolled at Unibo across the available covariates. The characteristics of the two samples are quite similar in most dimensions. The fact that 43 percent of the twins enroll in STEM while only 36.4 percent do so among non-twins is the most noticeable discrepancy. This may be due to the fact that, since sending two children to college simultaneously is costly, parents push them to choose more lucrative STEM majors. In terms of on-time graduation rates (0.53 vs. 0.52), grades (26.6 vs. 26.5), attendance to low-math high schools (0.34 vs. 0.37), and gender (0.59 vs. 0.50) the two distributions largely align. The fact that Emilia Romagna, where Unibo is located, is more represented among twins is likely due to the high burden parents would face were they to send twins to universities located in other regions. The same considerations are mirrored by the lower representation of twins from other regions.[9] [10]

To effectively implement TFE estimators, there needs to be adequate within-twin-pair variation in the treatment, that is, high school type. Table 2 shows that, within twin pairs, 30.7 percent (213 out of 693) display variation in the chosen type of high school.[11] There is also sufficient within-twin-pair variation for the outcomes of interest and the other twins' characteristics.[12] The percentages with a difference range from 38.9 (high average grade) to 33.7 (on-time graduation). For gender and high exit grade, there is variation within about 30 percent of the pairs.

Table 2: Within-twin-pair variations

	Number of pairs with within twin variation	Total number of pairs	Percentage with variation
STEM	246	693	35.498
On-Time Graduation	181	537	33.706
High Average Grade	227	583	38.937
Low Math	213	693	30.736
Female	197	693	28.427
High Exit Grade	194	693	27.994

3.2 Empirical Strategy

To establish a causal relationship between math exposure during high school and university outcomes, we use a TFE estimator. To illustrate our approach, we start with the baseline OLS

[9] Another parallel explanation of the observed differences by area of birth is that twin births are more common among older mothers, who are more prevalent in Emilia Romagna due to higher female labor market participation, which tends to delay childbirth. This effect is reinforced by the fact that older women are more likely to access fertility treatments, which increase the likelihood of twin births. See Pison and d'Addato (2006).

[10] For those regressors that may confound the effect of math exposure, a t-test of the difference between the two samples does not reveal any significant differences except for area of birth.

[11] From 1,396 individuals in the sample, we obtain 693 groups since we have 10 groups of triplets, adding up to 30 individuals. Subtracting these 30 from 1,396 and dividing by 2 we obtain 683 twin groups, to be added to the 10 triplet groups.

[12] To compute the variation in non-binary variables, we dichotomize the distribution of grades and the high school exit grade, whereas a high average grade is an average grade above or equal to the average and a high exit grade is a grade above or equal to 90 (out of 100).

estimator described, for each twin j ($j = 1, 2$), by:

$$Y_{jit} = \beta_0 + \beta_1 LowMath_{jit} + X_{it}\beta_2 + Z_{jit}\beta_3 + \epsilon_{ijt} \tag{1}$$

where Y_{jit} is the set of outcomes of interest for twin j in twin pair i, graduating from high school in year t; $LowMath_{jit}$ is a binary variable that takes value one if twin j in twin pair i completed a low-math high school in year t, and zero otherwise; X_{it} is a set of variables that vary by twin pair but not across twins within a pair (e.g., area of birth and year of graduation from high school); Z_{jit} is a set of variables that may vary across twins within a pair (e.g., gender, since we include fraternal twins, and high school exit grade); and ϵ_{ijt} is the error term.

The OLS estimate of the effect of a low-math background, represented in equation 1 by β_1, is generally biased due to observable or unobservable factors at the family level. The TFE estimator can be obtained by taking the first difference between equation 1 for twin 1 and the same equation for twin 2, in such a way that all characteristics that are common to twins within a pair are subsumed by the TFEs, to obtain:

$$Y_{1it} - Y_{2it} = \beta_1(LowMath_{1it} - LowMath_{2it}) + (Z_{1it} - Z_{2it})\beta_3 + (\epsilon_{1it} - \epsilon_{2it}) \tag{2}$$

In equation 2, all family effects are removed, so that the identification of the influence of math comes from an unbiased TFE estimate of β_1.[13] In other words, identification relies on the within-twin-pair variation in the choice of high school type. In terms of outcomes, we aim at assessing whether attending a low-math school affects students' choice to enroll in STEM fields and two dimensions of their university performance, namely, the likelihood of graduating on time, coded as a binary variable which equals one if a student graduates on time and zero otherwise, and grades.

Some issues need to be addressed. First, it is important to note that the TFE methodology was initially designed for application with monozygotic twin pairs (Behrman and Taubman, 1976; Griliches, 1979). However, our dataset does not include information on zygosity, which means that we cannot distinguish between identical (monozygotic) and fraternal (dizygotic) twins. As a result, our analysis includes fraternal twins, which introduces the potential for bias due to genetic differences that would otherwise be controlled for among identical twins. Despite this limitation, fixed effect methods similar to TFE have been applied to sibling pairs, starting as early as Gorseline (1939), with the aim of reducing the bias stemming from unobserved family background factors. Compared with sibling-based studies, our twin-based approach offers a significant advantage by reducing concerns about differential parental treatment. In sibling studies, parental behaviors—such as compensating for or reinforcing differences between children—can introduce confounding effects, complicating causal interpretations. However, as highlighted by Bhalotra and Clarke (2023), by focusing on twins

[13] An equivalent formulation of equation 2, which rather than first differences emphasizes the term τ_j capturing TFEs, can be written as: $Y_{jit} = \beta_0 + \beta_1 LowMath_{jit} + Z_{jit}\beta_3 + \tau_j + \epsilon_{jit}$.

we minimize these potential sources of bias, since parents are generally more inclined to treat twins alike, given their shared developmental stage and simultaneous milestones. This uniformity in parental treatment among twins enhances the validity of causal inferences drawn from twin-based analyses, even when zygozity info is lacking.

Another challenge in twin analysis is the potential for measurement error in the primary independent variable, which can be amplified by the TFE estimator.[14] However, in our specific context, the likelihood of measurement error in high school background variables is minimal or entirely absent, thanks to the accuracy and reliability of our administrative data sources.

A potential weakness of twin analysis is that it may be underpowered, as identification relies on differences between twins that may be small or subtle. This is often the case with variables such as birth weight or years of education (Bhalotra and Clarke, 2023). However, in our study, this concern is mitigated by the binary nature of the treatment. This clear-cut variation in educational exposure strengthens the power of our approach, allowing a meaningful identification of the effects of high school math exposure on later academic outcomes.

Another potential source of bias in twin studies stems from their reliance on twin registries, which often depend on voluntary participation or active recruitment by researchers. This process can introduce selection biases, such as an overrepresentation of identical twins, which can lead to inflated estimates of heritability (Webbink et al., 2006). In contrast, our data collection includes all twins in the target population using administrative records, ensuring comprehensive coverage and eliminating concerns about volunteer or recruitment bias.

Although the above considerations primarily address the internal validity of the results, external validity poses additional challenges. In sibling analyses, it has been argued that introducing fixed effects with a binary treatment can induce selection into the identifying sample, systematically skewing representation across families. This results in estimates derived from a limited subset of the population, often overemphasizing larger families (Miller et al., 2023). However, our twin-based approach avoids this pitfall since, unlike subsequent births, twin births are typically unplanned, thus reducing the risk of such selection bias.[15]

4 Results

Figure 1 (and Table B.1 in Appendix B) present the results of TFE regressions conducted on our sample of twins, focusing on three key outcomes: the likelihood of enrolling in a STEM degree, the likelihood of completing a degree on time, and standardized exam grades. The treatment, low math, is a binary indicator equal to one for students who attended a high school with a low-math curriculum, and zero otherwise. We also control for gender, which is

[14]To mitigate this issue, Ashenfelter and Krueger (1994) proposed an IV methodology, where the instrument is constructed using cross-reported information between twins. However, this type of data is not available in our dataset.

[15]A notable exception may occur in cases involving fertility treatments, which could influence the likelihood of twin births.

the only available exogenous variable that varies within twin pairs.[16]

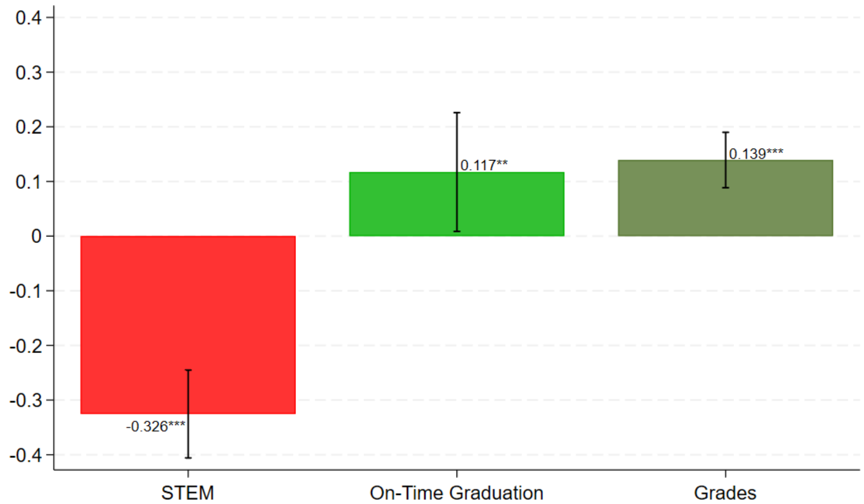

Figure 1: The effect of low math in high school on student outcomes: TFE estimates

Notes: Point estimates and corresponding 95 percent CIs obtained from TFE regressions of the outcome variables on Low Math, controlling for gender. The unit of observation is a student for STEM and On-Time graduation and a grade for Grades. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math school type. Female is a binary variable taking value one if a student is female.

These findings highlight a clear negative impact of attending a low-math high school on the decision to pursue a STEM degree. Specifically, students from low-math schools are 32.6 percentage points less likely to enroll in STEM programs, a result that is both sizable and highly statistically significant. In contrast, when examining academic performance, the results reveal a notable positive effect of a low-math curriculum. Students from low-math schools are 11.7 percentage points more likely to graduate on time and achieve a 0.139 standard deviation improvement in grades. Our results also align with well-documented gender dynamics in STEM fields (Sax et al., 2015; Ellis et al., 2016), since being female is associated with a reduced likelihood of pursuing STEM, while showing a positive (though not statistically significant) association with on-time graduation and a significant boost in grades. The positive impact of a low-math high school background on university performance contradicts much of the existing literature (Ely and Hittle, 1990), and is therefore puzzling.[17]

To corroborate our findings, in Appendix B we perform a number of robustness checks,

[16]To be noticed is that this variable only captures differences within opposite-sex pairs, since within same-sex pairs it is absorbed by the TFE.

[17]One possible explanation is that this effect may be driven by students who attended a specific track within the low-math category—the classical *lyceum*—that often attracts high-achieving students. However, this track includes relatively few students. Among the 213 twin pairs attending different school types, only for 72 the low-math school of choice is the classical *lyceum*. Controlling for the latter, or dropping from the sample these 72 pairs, does not alter the positive sign of the coefficients, even though the one on on-time graduation is no longer significant (results not reported for brevity).

comparisons, and extensions. Specifically, in Table B.2 we show that the positive influence of low math on performance persists even after accounting for a composition effect, where treated students may opt for less demanding non-STEM fields.[18] In Tables B.3-B.5 we compare TFE estimates and OLS and GLS [19] estimates obtained using the twins sample, as well as OLS estimates obtained using the entire Unibo sample. Especially for the indicators of performance, the alternative estimates reveal a downward bias in the effect of low math.[20] Furthermore, by examining specific dimensions of heterogeneity, Appendix C highlights the differential impact of math exposure accounting for gender, zygosity, ability, and major choice.

Next, to better isolate the impact of math exposure on our key outcomes, we examine a 2010 high school reform that significantly increased math content in traditionally low-math curricula. This reform, which became effective for students graduating from high school and entering university in 2015, provides yet another natural experiment, alongside the occurrence of twin births, within our dataset.

Table 3 presents TFE estimates for the pre-reform and post-reform cohorts separately. A comparison of columns 1 and 2 reveals that the negative impact of attending a low-math high school on STEM enrollment persists across both periods but becomes markedly stronger—an additional 16.8 percentage points—following the reform. A t-test for the pre/post difference in the effect of low math confirms that the post-reform decline is a statistically significant one (as shown by the p-value reported at the bottom of the table). Hence, contrary to expectations, increasing math exposure in low-math high schools does not spark greater interest in math-intensive university degrees; instead, it further discourages students from pursuing STEM fields.

Turning to academic performance, we discover that the impact of a traditionally low-math background is actually a post-reform occurrence, as no significant effects are observed in the pre-reform period (columns 3 and 5). The p-values confirm that the a pre/post difference is detectable for grades but not for on-time graduation.[21] This finding sheds light on the seemingly counterintuitive result in Figure 1 (and Table B.1), where the positive influence of a low-math background on academic performance appears to span the entire sample period, ultimately aligning our findings with the existing literature (Joensen and Nielsen, 2016; De Philippis, 2023).[22]

[18]In the same table we show that our results are robust to controlling for ability, as proxied by the high school exit grade.

[19]The GLS approach has been used in twin studies (Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998) because it can account for the correlation in error terms between twins.

[20]We also show that the results in Table B.1 are robust to restricting the sample to the pre-Covid period (Table B.6), despite the fact that the pandemic might have affected students in several dimensions (Rodríguez-Planas, 2022), and to bootstrapping the standard errors (Table B.7).

[21]This result holds consistent whether or not STEM enrollment is included as a control variable, as demonstrated in Appendix B (Table B.8).

[22]Rather than splitting the sample between the pre- and post-reform periods, as we do in Table 3, identical results (unreported for brevity) can be obtained using the entire sample and adding an interaction term involving low math and a binary variable taking value one for the post-reform period (and zero for the pre-reform period).

Table 3: The effect of the high school reform on student outcomes: TFE estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	STEM		On-Time Graduation		Grades	
	Pre	Post	Pre	Post	Pre	Post
Low Math	-0.213***	-0.381***	0.045	0.179**	0.056	0.201***
	(0.066)	(0.052)	(0.083)	(0.075)	(0.042)	(0.033)
Female	-0.144**	-0.149***	0.100	0.051	0.289***	0.148***
	(0.072)	(0.057)	(0.085)	(0.093)	(0.044)	(0.038)
Constant	0.588***	0.649***	0.388***	0.497***	-0.224***	-0.136***
	(0.052)	(0.034)	(0.052)	(0.050)	(0.029)	(0.024)
R-squared	0.335	0.404	0.244	0.382	0.245	0.247
Observations	478	918	478	614	8196	11895
Twins	237	456	237	300	237	456
Pre/Post Difference (*p*-value)	0.046		0.231		0.007	

Notes: The unit of observation is a student in columns 1-4 and a grade in columns 5-6. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Robust standard errors in parentheses: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

While the reform presents an opportunity for an alternative identification strategy based on a DiD estimator, the latter cannot exploit TFEs for identification, since it relies on pre- and post-reform variation that is absent for twins who enroll at university in the same year. For comparison sake, Appendix D contains a DiD analysis on the entire sample of Unibo students, leading to a much weaker impact of the reform on STEM enrollment and a nil or even reversed effect on performance. These differences are attributable to the selection bias that potentially affects DiD, stemming from non-random selection into treatment (Ashenfelter and Card, 1985), to which the TFE approach is immune.

In conclusion, our TFE estimates reveal that a low-math high school background significantly reduces the likelihood of enrolling in STEM programs, an effect that intensifies after a reform aimed at increasing the math content, as the reform further discouraged treated students from pursuing STEM majors. However, the introduction of additional math content due to the reform led to notable improvements in the academic performance of treated students. These gains persist even after accounting for a composition effect, making treated students opt for non-STEM fields. Together, these findings suggest that while enhanced math exposure does not make STEM majors more attractive, it equips students with the skills necessary to thrive in their university studies, regardless of their chosen field.

5 Discussion and Conclusion

Our study provides evidence of the profound impact of high school mathematics exposure on college major choices and academic performance, leveraging a unique dataset of twins from

Lastly, Table B.9 shows that the results in Table 3 are broadly robust to restricting the post-reform sample to the pre-Covid period.

the University of Bologna. Employing a TFE estimator, we were able to control for unobserved family background factors, shedding light on how math exposure shapes educational trajectories. Our findings enrich the existing literature by underscoring the importance of high school math in pursuing STEM majors and achieving broader academic success.

One striking insight is the sharp reduction in STEM enrollment—by 32.6 percentage points—among students from low-math high schools compared to their high-math counterparts. Paradoxically, during the entire period under examination, these students excelled academically, demonstrating higher on-time graduation rates and grades. This suggests that while rigorous math preparation is essential for choosing STEM fields, less intensive math exposure may foster success in the entire academic domain. However, the evaluation of a national high school reform that added math content to traditionally low-math curricula allows us to shed further light on the underlying dynamics. On the one hand, the reform further reduced STEM enrollment, suggesting that more exposure to math in high school does not make a STEM career more attractive, at least among those students who had chosen a low-math track to begin with. In other words, taking more math within a traditionally low-math curriculum does not sort the same effect as choosing a high-math curriculum in the first place. On the other hand, splitting the sample between the pre- and post-reform periods allows us to highlight that the superior performance of students from traditionally low-math schools manifests itself only after the reform, that is, it is due to the added math content to their curricula, despite the fact that it remained lower than in traditionally high-math schools. This dual effect illustrates that increasing math exposure, while not necessarily fostering STEM interest, equips students with skills that bolster their broader academic success.

Our analysis underscores the unique value of twin-based research in uncovering the nuanced effects of educational policies. By controlling for unobserved variables and eluding selection issues, such methodologies can provide clearer insight and inform more effective policy decisions. Future research and policymaking would benefit from incorporating such approaches to better capture the complexities of educational outcomes. Specifically, our findings carry important policy implications, as they suggest that educational reforms must balance the need for rigorous math education with the promotion of student achievement in diverse fields. Designing flexible curricula that build foundational math skills while accommodating varied interests could better support student outcomes.

References

Altonji, J. G. (1995). The Effects Of High School Curriculum On Education And Labor Market Outcomes. The Journal of Human Resources, 30(3):409–438.

Ashenfelter, O. and Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. The Review of Economics and Statistics, 67(4):648–660.

Ashenfelter, O. and Krueger, A. (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. The American Economic Review, 84(5):1157–1173.

Ashenfelter, O. and Rouse, C. (1998). Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins. The Quarterly Journal of Economics, 113(1):253–284.

Aughinbaugh, A. (2012). The Effects of High School Math Curriculum on College Attendance: Evidence From the NLSY97. Economics of Education Review, 31(6):861–870.

Behrman, J. and Taubman, P. (1976). Intergenerational Transmission of Income and Wealth. The American Economic Review, 66(2):436–440.

Behrman, J. R. and Rosenzweig, M. R. (1999). "Ability" Biases in Schooling Returns and Twins: A Test and New Estimates. Economics of Education Review, 18(2):159–167.

Behrman, J. R., Xiong, Y., and Zhang, J. (2015). Cross-Sectional Schooling-Health Associations Misrepresented Causal Schooling Effects on Adult Health and Health-Related Behaviors: Evidence From the Chinese Adults Twins Survey. Social Science & Medicine, 127:190–197.

Bertocchi, G., Bonacini, L., and Murat, M. (2023). Adams and Eves: High School Math and the Gender Gap in Economics Majors. Economic Inquiry, 61(4):798–817.

Bhalotra, S. and Clarke, D. (2023). Analysis of Twins. In Handbook of Labor, Human Resources and Population Economics, pages 1–37. Springer Nature.

De Philippis, M. (2023). STEM graduates and secondary school curriculum: does early exposure to science matter? Journal of Human Resources, 58(6):1914–1947.

Delaney, J. M. and Devereux, P. J. (2019). Understanding Gender Differences in STEM: Evidence From College Applications. Economics of Education Review, 72:219–238.

Delaney, J. M. and Devereux, P. J. (2025). Gender Differences in Graduate Degree Choices. Journal of Economic Behavior & Organization, 230:106882.

Ellis, J., Fosdick, B. K., and Rasmussen, C. (2016). Women 1.5 Times More Likely To Leave STEM Pipeline After Calculus Compared To Men: Lack Of Mathematical Confidence A Potential Culprit. PLOS ONE, 11(7):e0157447.

Ely, D. P. and Hittle, L. (1990). The Impact of Math Background on Performance in Managerial Economics and Basic Finance Courses. Journal of Financial Education, 19:59–61.

Feigenbaum, J. J. and Tan, H. R. (2020). The Return to Education in the Mid-Twentieth Century: Evidence From Twins. The Journal of Economic History, 80(4):1101–1142.

Gorseline, D. (1939). The Impact of Family on Educational Achievement. Journal of Educational Research, 32:193–202.

Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey. Journal of Political Economy, 87(5):S37–S64.

Joensen, J. S. and Nielsen, H. S. (2009). Is There a Causal Effect of High School Math on Labor Market Outcomes? Journal of Human Resources, 44(1):171–198.

Joensen, J. S. and Nielsen, H. S. (2016). Mathematics and Gender: Heterogeneity in Causes and Consequences. Economic Journal, 126:1129–1163.

Levine, P. B. and Zimmerman, D. J. (1995). The Benefit of Additional High School Math and Science Classes for Young Men and Women. Journal of Business and Economic Statistics, 13(2):137–149.

Li, H., Liu, P. W., and Zhang, J. (2012). Estimating Returns to Education Using Twins in Urban China. Journal of Development Economics, 97(2):494–504.

Lin, M. J. and Liu, J. T. (2009). Do Lower Birth Weight Babies Have Lower Grades? Twin Fixed Effect and Instrumental Variable Method Evidence From Taiwan. Social Science & Medicine, 68(10):1780–1787.

Long, M. C., Iatarola, P., and Conger, D. (2009). Explaining Gaps In Readiness For College-Level Math: The Role Of High School Courses. Education Finance and Policy, 4(1):1–33.

Miller, D. L., Shenhav, N. A., and Grosz, M. (2023). Selection into Identification in Fixed Effects Models, With Application to Head Start. Journal of Human Resources, 58(5):1523–1566.

Pison, G. and d'Addato, A. V. (2006). Frequency of Twin Births in Developed Countries. Twin Research and Human Genetics, 9(2):250–259.

Rodríguez-Planas, N. (2022). COVID-19, College Academic Performance, and the Flexible Grading Policy: A Longitudinal Analysis. Journal of Public Economics, 207:104606.

Rose, H. and Betts, J. R. (2004). The Effect of High School Courses on Earnings. Review of Economics and Statistics, 86(2):497–513.

Rosenzweig, M. R. and Zhang, J. (2009). Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy. The Review of Economic Studies, 76(3):1149–1174.

Rosenzweig, M. R. and Zhang, J. (2013). Economic Growth, Comparative Advantage, and Gender Differences in Schooling Outcomes: Evidence From the Birthweight Differences of Chinese Twins. Journal of Development Economics, 104:245–260.

Sax, L. J., Kanny, M. A., Riggers-Piehl, T. A., Whang, H., and Paulson, L. (2015). Women In STEM: A Gender Gap To Innovation. Economics of Education Review, 49:186–197.

Webbink, D., Roeleveld, J., and Visscher, P. M. (2006). Identification of Twin Pairs From Large Population-Based Samples. Twin Research and Human Genetics, 9(4):496–500.

Zhang, J., Liu, P. W., and Yung, L. (2007). The Cultural Revolution and Returns to Schooling in China: Estimates Based on Twins. Journal of Development Economics, 84(2):631–639.

ONLINE APPENDICES FOR
Math Exposure and University Performance: Causal Evidence from Twins

Graziella Bertocchi, Luca Bonacini, Majlinda Joxhe, Giuseppe Pignataro

A Italy vs. the University of Bologna

Figure A.1 highlights the similarity between the geographic distribution of students across Italy (map on the left) and those enrolled at the University of Bologna (Unibo, map on the right). This alignment underscores the representativeness of Unibo as a reflection of the broader Italian student population, despite slight regional variations. One example is the higher proportion of Unibo students from the *Marche* region (highlighted in dark blue) compared with the national average, but this can be attributed to the proximity of *Marche* to *Emilia Romagna*, where Unibo is located. Interestingly, the overall distribution of students between northern and southern Italy remains comparable, despite the presence of metropolitan hubs like Milan, Turin, Rome, and Naples, which host a denser concentration of universities. This geographical spread reinforces the position of Unibo as a key academic institution that attracts students from all parts of Italy.

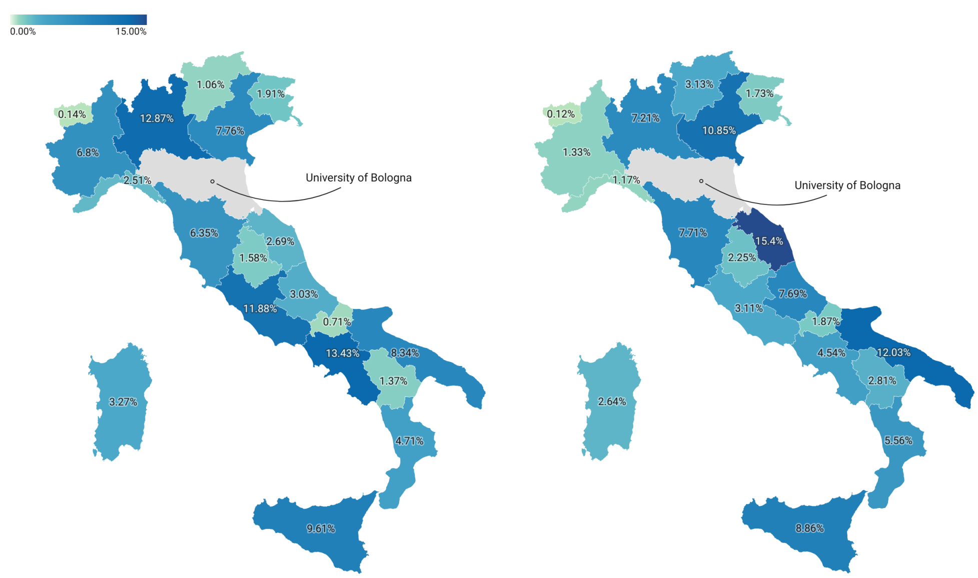

Figure A.1: Shares of students enrolled at university from 2010 to 2021, by region of residence

Notes: A darker shade indicates a larger share. On the left, students enrolled at all Italian universities, on the right, students enrolled at Unibo. The data source is *Anagrafe Nazionale degli Studenti*.

Figure A.2 provides further evidence by focusing on high school background. During the past decade, the distributions by high school types—low math vs. high math—are quite similar for students from all over Italy (left) and those from Unibo (right). Specifically, 31.47 percent of the students enrolled in all Italian universities come from low-math schools. At Unibo, this proportion is 36.98. This consistency underscores that Unibo's student population mirrors quite closely the characteristics of students at other Italian universities, even when segmented by educational pathways, reinforcing its representativeness on a national scale.

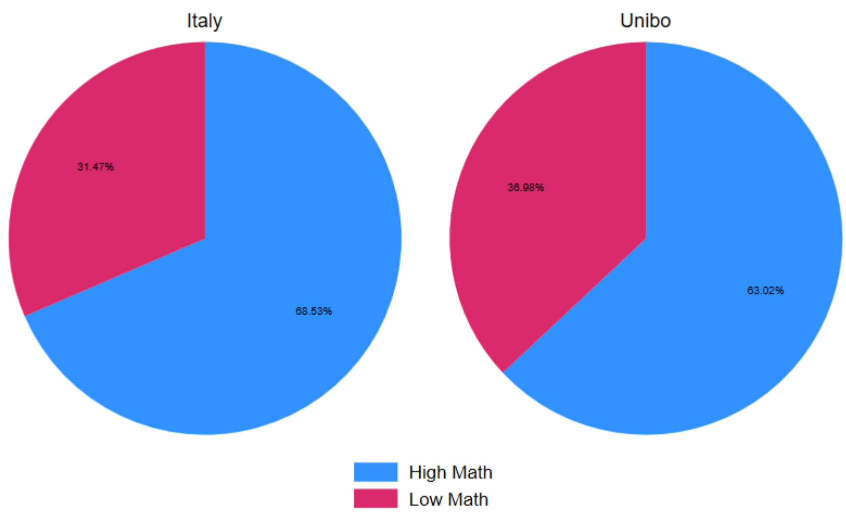

Figure A.2: Shares of students enrolled at university from 2010 to 2021, by high school type

Notes: On the left, students enrolled at all Italian universities, On the right, students enrolled at Unibo. The data source is *Anagrafe Nazionale degli Studenti*.

B Robustness Checks

As Figure 1, Table B.1 presents the results of TFE regressions conducted on the sample of twins. The rest of the appendix delves into a series of robustness checks to validate and deepen our understanding of the main findings of the paper. For Tables B.2-B.7, the term of comparison is Table B.1, while for Tables B.8 and B.9 the term of comparison is Table 3.

Table B.2 extends the TFE estimates in Table B.1 by adding two additional controls among those that can vary between twins within a pair. We begin with the high school exit grade, which serves as a proxy for academic ability (a passing grade is 60 out of 100). Although we acknowledge the potential endogeneity of this variable—which makes it unsuitable for inclusion in the preferred specifications presented in Table B.1—we use it here to explore its influence. Consistent with expectations, students with higher ability are more likely to pursue STEM majors (column 1), graduate on time (column 3), and achieve higher grades (column 6). However, incorporating this ability proxy does not diminish the effect of low math on any outcome. In fact, the impact of low math becomes even more pronounced,

Table B.1: The effect of high school math on student outcomes: TFE estimates

	(1) STEM	(2) On-Time Graduation	(3) Grades
Low Math	-0.326***	0.117**	0.139***
	(0.041)	(0.055)	(0.026)
Female	-0.151***	0.080	0.217***
	(0.045)	(0.063)	(0.029)
Constant	0.631***	0.446***	-0.177***
	(0.028)	(0.036)	(0.018)
R-squared	0.378	0.328	0.246
Observations	1396	1092	20091
Twins	693	537	693

Notes: The unit of observation is a student in columns 1 and 2 and a grade in column 3. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

particularly for on-time graduation and grades.[1] The second regressor we introduce in Table B.2 is STEM enrollment (with and without the exit grade). Since STEM curricula are generally more demanding, it is plausible that the observed impact of high school math on performance might be partially explained by a composition effect. In other words, students from low-math high schools may avoid STEM majors, thereby influencing performance metrics. However, controlling for STEM enrollment reveals that while the effect of math is somewhat attenuated, it remains substantial within twin pairs where one twin is enrolled in STEM and the other in a different field (columns 2 and 5).

Table B.2: The effect of high school math on student outcomes: TFE estimates with additional controls

	(1) STEM	(2) On-Time Graduation	(3) On-Time Graduation	(4) On-Time Graduation	(5) Grades	(6) Grades	(7) Grades
Low Math	-0.327***	0.108*	0.122**	0.106*	0.095***	0.170***	0.106***
	(0.041)	(0.057)	(0.055)	(0.057)	(0.027)	(0.025)	(0.026)
Female	-0.167***	0.074	0.059	0.049	0.192***	0.125***	0.082***
	(0.045)	(0.064)	(0.062)	(0.063)	(0.029)	(0.028)	(0.028)
Exit Grade	0.006***		0.008***	0.009***		0.025***	0.026***
	(0.002)		(0.002)	(0.002)		(0.001)	(0.001)
STEM		-0.032		-0.055	-0.144***		-0.217***
		(0.053)		(0.052)	(0.025)		(0.025)
Constant	0.144	0.466***	-0.251	-0.241	-0.085***	-2.271***	-2.250***
	(0.164)	(0.049)	(0.195)	(0.195)	(0.024)	(0.104)	(0.104)
R-squared	0.386	0.328	0.347	0.348	0.248	0.265	0.268
Observations	1396	1092	1092	1092	20091	20091	20091
Twins	693	537	537	537	693	693	693

Notes: The unit of observation is a student in columns 1-4 and a grade in columns 5-7. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Exit Grade is the high school exit grade. Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Tables B.3-B.5, separately for each of the three outcomes of interest, for comparison purposes we present alternative estimators and samples, other than the TFE estimates on the twins sample in Table B.1. Table B.3 focuses on STEM enrollment and presents results

[1] Notably, controlling simultaneously for both STEM enrollment and high school ability does not alter the results (columns 4 and 7).

Table B.3: The effect of high school math on STEM enrollment: OLS and GLS estimates

	(1)	(2)	(3)
		STEM	
	OLS	GLS	OLS
Low Math	-0.310***	-0.329***	-0.297***
	(0.028)	(0.028)	(0.003)
Female	-0.098***	-0.083***	-0.174***
	(0.030)	(0.029)	(0.003)
Constant	0.648***	0.649***	0.626***
	(0.053)	(0.052)	(0.005)
High School Graduation Year	Yes	Yes	Yes
Area of Birth	Yes	Yes	Yes
R-squared	0.128	0.155	0.170
Observations	1396	1396	123015
Twins	693	693	

Notes: The unit of observation is a student. STEM is a binary variable taking value one if a student enrolls in a STEM degree. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Robust standard errors clustered at the twin-pair level in columns 1 and 2 and robust standard errors in column 3 in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

from three specifications: OLS estimates in column 1, GLS estimates in column 2, and OLS estimates on the population of all Unibo students in column 3. All models control for gender, high school graduation year, and area of birth, with standard errors clustered at the twin-pair level, except for the OLS using the full Unibo population where we use robust standard errors. Across all specifications, the negative effect of a low-math background on STEM enrollment is consistent with the TFE estimates.

Table B.4: The effect of high school math on on-time graduation: OLS and GLS estimates

	(1)	(2)	(3)	(4)	(5)	(6)
			On-Time Graduation			
	OLS	OLS	GLS	GLS	OLS	OLS
Low Math	0.045	0.031	0.050	0.036	0.020***	-0.019***
	(0.037)	(0.038)	(0.037)	(0.038)	(0.004)	(0.004)
Female	0.086**	0.080**	0.091**	0.085**	0.095***	0.072***
	(0.037)	(0.038)	(0.038)	(0.038)	(0.003)	(0.004)
STEM		-0.049		-0.050		-0.131***
		(0.035)		(0.035)		(0.004)
Constant	0.371***	0.403***	0.364***	0.398***	0.395***	0.477***
	(0.054)	(0.058)	(0.054)	(0.059)	(0.005)	(0.006)
High School Graduation Year	Yes	Yes	Yes	Yes	Yes	Yes
Area of Birth	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.025	0.026	0.031	0.032	0.020	0.033
Observations	1092	1092	1092	1092	92637	92637
Twins	537	537	537	537		

Notes: The unit of observation is a student. On-Time Graduation is a binary variable taking the value of one if a student graduates on time. Low Math is a binary variable taking the value of one if a student attended a low-math high school type. Female is a binary variable taking the value of one if a student is female. Robust standard errors clustered at the twin-pair level in columns 1 and 2 and robust standard errors in column 3 in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.4 presents analogous results for on-time graduation, initially without and then with STEM enrollment as an additional control. In all specifications, the coefficient for the low-math variable is smaller and less significant compared with Table B.1, especially in columns 1-4 based on the twins sample, pointing to a downward bias in the alternative estimators used in the table.[2] A similar pattern emerges for grades in Table B.5: a low-math background exerts no effect in the OLS and GLS based on the twins sample, while the OLS based on the full sample retain a positive coefficient, even after controlling for STEM, but in both cases the coefficients are smaller than in the TFE. Overall, these comparisons demonstrate that factors related to family background, which are not captured by OLS estimates, contribute substantially to the effect of high-school math on subsequent university careers, particularly for the two measures of performance.

Table B.5: The effect of high school math on grades: OLS and GLS estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Grades					
	OLS	OLS	GLS	GLS	OLS	OLS
Low Math	0.023	-0.008	0.041	0.009	0.117***	0.058***
	(0.040)	(0.041)	(0.039)	(0.041)	(0.002)	(0.002)
Female	0.204***	0.189***	0.189***	0.172***	0.147***	0.113***
	(0.043)	(0.042)	(0.042)	(0.042)	(0.002)	(0.002)
STEM		-0.112***		-0.121***		-0.198***
		(0.041)		(0.042)		(0.002)
Constant	-0.247***	-0.175**	-0.169**	-0.096	-0.206***	-0.083***
	(0.070)	(0.074)	(0.074)	(0.078)	(0.003)	(0.003)
High School Graduation Year	Yes	Yes	Yes	Yes	Yes	Yes
Area of Birth	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.019	0.021	0.018	0.021	0.016	0.024
Observations	22071	22071	22071	22071	1659054	1659054
Twins	817	817	817	817		

Notes: The unit of observation is a grade. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking the value of one if a student attended a low-math high school type. Female is a binary variable taking the value of one if a student is female. Robust standard errors clustered at the twin-pair level in columns 1 and 2 and robust standard errors in column 3 in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.6: The effect of high school math on student outcomes: The pre-Covid sample

	(1)	(2)	(3)
	STEM	On-Time Graduation	Grades
Low Math	-0.287***	0.056	0.100***
	(0.045)	(0.080)	(0.031)
Female	-0.161***	0.080	0.291***
	(0.049)	(0.088)	(0.037)
Constant	0.622***	0.430***	-0.208***
	(0.032)	(0.052)	(0.023)
R-squared	0.382	0.250	0.258
Observations	1147	562	13690
Twins	570	264	535

Notes: The unit of observation is a student in columns 1 and 2 and a grade in column 3. STEM is a binary variable taking the value of one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking the value of one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking the value of one if a student attended a low-math high school type. Female is a binary variable taking the value of one if a student is female. Robust standard errors are reported in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.6 presents a robustness check to account for the potential influence of the Covid-19 pandemic on students' choices and performance. Recognizing that the pandemic may have significantly altered educational dynamics, we confine our analysis to the pre-Covid period,

[2]For the OLS run over the full sample of students, the coefficient on the low-math variable becomes even negative when controlling for STEM (column 6).

restricting the sample to observations from the beginning of the 2019/2020 academic year and earlier. The results are largely consistent with those reported in Table B.1 as far as STEM enrollment and grades are concerned, while the influence of low math on on-time graduation remains positive but no longer statistically significant.

Table B.7: The effect of high school math on student outcomes: Bootstrapping standard errors

	(1) STEM	(2) On-Time Graduation	(3) Grades
Low Math	-0.326***	0.117	0.139***
	(0.054)	(0.073)	(0.027)
Female	-0.151**	0.080	0.217***
	(0.061)	(0.082)	(0.029)
Constant	0.631***	0.446***	-0.177***
	(0.038)	(0.046)	(0.019)
R-squared	0.378	0.328	0.246
Observations	1396	1092	20091
Twins	693	537	693

Notes: The unit of observation is a student in columns 1 and 2 and a grade in column 3. STEM is a binary variable taking the value of one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking the value of one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking the value of one if a student attended a low-math high school type. Female is a binary variable taking the value of one if a student is female. Bootstrapped standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table B.7 presents an additional robustness check, where the standard errors of the coefficients on the low-math variable are computed via a bootstrap procedure— a robust statistical method that provides reliable inference by resampling the dataset to generate multiple simulated samples.[3] This ensures that the confidence intervals and the p-values reflect the structure of the data, rather than rely solely on parametric assumptions. This approach allows to estimate the variability of coefficients and test their significance with greater precision, particularly when working with smaller samples or when the underlying assumptions of traditional parametric tests might be challenged. Previous results are confirmed for STEM enrollment and grades, while we detect a loss of significance for on-time graduation.

The last two tables extend the analysis of the high school reform in Table 3. Table B.8 shows the impact of the reform on both dimensions of performance is robust to adding STEM enrollment as a control variable. Lastly, Table B.9 examines the post-reform impact of the high school math reform within the pre-Covid sample. The results from Table 3 are confirmed, despite a loss of significance of the effect on on-time graduation once STEM enrollment is included (column 3).

[3]The bootstrap procedure employed here involves resampling with 2,000 replications.

Table B.8: The effect of the high school math reform: Controlling for STEM enrollment

	(1)	(2)	(3)	(4)
	On-Time Graduation		Grades	
	Pre	Post	Pre	Post
Low Math	0.019	0.204**	0.010	0.175***
	(0.084)	(0.079)	(0.042)	(0.035)
Female	0.082	0.064	0.249***	0.137***
	(0.085)	(0.095)	(0.044)	(0.038)
STEM	-0.123*	0.067	-0.222***	-0.071**
	(0.071)	(0.077)	(0.037)	(0.035)
Constant	0.460***	0.451***	-0.091***	-0.090***
	(0.064)	(0.074)	(0.035)	(0.033)
R-squared	0.251	0.383	0.249	0.247
Observations	478	614	8196	11895
Twins	237	300	237	456

Notes: The unit of observation is a student in columns 1-2 and a grade in columns 3-4. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high-school type. Female is a binary variable taking value one if a student is female. STEM is a binary variable taking value one if a student enrolls in a STEM degree. Robust standard errors in parentheses: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table B.9: The post-reform effect of high school math: The pre-Covid sample

	(1)	(2)	(3)	(4)	(5)
	STEM	On-Time Graduation	On-Time Graduation	Grades	Grades
Low Math	-0.337***	0.391**	0.297	0.140***	0.149***
	(0.062)	(0.165)	(0.197)	(0.046)	(0.048)
Female	-0.169**	-0.217	-0.268	0.306***	0.311***
	(0.066)	(0.226)	(0.235)	(0.065)	(0.066)
STEM			-0.185		0.032
			(0.253)		(0.051)
Constant	0.643***	0.581***	0.723***	-0.235***	-0.256***
	(0.040)	(0.116)	(0.210)	(0.039)	(0.051)
R-squared	0.417	0.287	0.287	0.276	0.276
Observations	669	121	121	5726	5726
Twins	333	54	54	303	303

Notes: The unit of observation is a student in columns 1-3 and a grade in columns 4-5. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Robust standard errors in parentheses: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

C Heterogeneity Analysis

To explore the potential heterogeneity in the effect of high school math, in this appendix we replicate the TFE estimates across different groups within the twin population, focusing on gender, zygosity, ability, and major choice. However, these results should be interpreted with caution due to the reduced sample size, which may limit the precision and robustness of the findings.

In Table C.1, we look at STEM enrollment. Columns 1 and 2 restrict the sample to twin pairs of the same sex, separately for males and females, revealing no substantial sex-related differences in the impact of a low-math background. When we focus on twin pairs of the same sex in column 3, the coefficient size remains consistent with the previous two columns. To be noticed is that these first three samples include all identical twins as well as some fraternal twins. In column 4, the sample is restricted to opposite-sex twin pairs, a group exclusively comprising fraternal twins (albeit not all of them), which represent 28 percent of the twin

Table C.1: The effect of high school math on STEM enrollment: Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
	STEM					
	Males	Females	Same Sex	Opposite Sex	Low Exit Grade	High Exit Grade
Low Math	-0.286***	-0.296***	-0.294***	-0.385***	-0.358***	-0.407***
	(0.099)	(0.056)	(0.049)	(0.074)	(0.055)	(0.106)
Female				-0.133***	-0.108*	-0.216**
				(0.049)	(0.061)	(0.100)
Constant	0.583***	0.508***	0.536***	0.615***	0.570***	0.748***
	(0.026)	(0.029)	(0.020)	(0.034)	(0.037)	(0.064)
R-squared	0.388	0.435	0.432	0.241	0.405	0.415
Observations	375	623	998	398	681	324
Twins	187	309	496	197	338	161

Notes: The unit of observation is a student. STEM is a binary variable taking value one if a student enrolls in a STEM degree. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Low (high) exit grade is a high school exit grade below (above) 90 (out of 100). Robust standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

pairs. Here, the absolute value of the coefficient rises sharply, indicating a significantly larger negative effect of low math, reaching 38.5 percentage points. Interestingly, while female twin pairs are affected by low math as much as male twin pairs, being in an opposite-sex twin pair implies a lower likelihood of majoring in STEM. When examining the impact of high school math across different ability levels, the effect is stronger for high-ability students, although it remains evident for their low-ability counterparts (columns 5 and 6).[4]

Table C.2: The effect of high school math on on-time graduation: Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	On-Time Graduation							
	Males	Females	Same Sex	Opposite Sex	Low Exit Grade	High Exit Grade	Non-STEM	STEM
Low Math	-0.100	0.137*	0.086	0.173*	0.122	0.154	0.148*	0.005
	(0.121)	(0.075)	(0.065)	(0.102)	(0.082)	(0.110)	(0.086)	(0.115)
Female				0.062	0.107	0.044	0.212**	-0.186
				(0.074)	(0.095)	(0.131)	(0.101)	(0.128)
Constant	0.505***	0.508***	0.509***	0.428***	0.355***	0.629***	0.341***	0.588***
	(0.029)	(0.037)	(0.025)	(0.040)	(0.050)	(0.080)	(0.074)	(0.062)
R-squared	0.484	0.365	0.410	0.109	0.283	0.461	0.375	0.511
Observations	297	496	793	299	574	233	430	295
Twins	147	242	389	148	283	112	212	141

Notes: The unit of observation is a student. On-Time Graduation is a binary variable taking value one if a student graduates on time. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Low (high) exit grade is a high school exit grade below (above) 90 (out of 100). Robust standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Heterogeneities are less pronounced for on-time graduation (Table C.2), possibly due to the limited sample size in the sub-samples. For grades (Table C.3), the positive influence of low math is driven by the female twin pairs (column 2) and by non-STEM students (column 7).

[4]It is important to note that adopting the TFE methods ensures that, within each sub-sample, the coefficient on low math reflects differences within twin pairs sharing the same ability level. This is why the coefficients on low math in both sub-samples can be—and indeed are—larger than those derived without dividing the sample along the ability dimension.

Table C.3: The effect of high school math on grades: Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Grades			
	Males	Females	Same Sex	Opposite Sex	Low Exit Grade	High Exit Grade	Non-STEM	STEM
Low Math	0.092	0.180***	0.164***	0.086*	0.166***	0.183***	0.154***	-0.065
	(0.081)	(0.032)	(0.030)	(0.050)	(0.041)	(0.049)	(0.040)	(0.059)
Female				0.232***	0.102**	0.270***	0.138***	0.131**
				(0.032)	(0.049)	(0.056)	(0.046)	(0.066)
Constant	-0.140***	0.015	-0.043***	-0.183***	-0.360***	0.119***	-0.122***	-0.120***
	(0.019)	(0.017)	(0.013)	(0.022)	(0.030)	(0.038)	(0.037)	(0.034)
R-squared	0.272	0.252	0.268	0.189	0.205	0.220	0.261	0.289
Observations	5382	9098	14480	5611	9112	5099	7622	5203
Twins	187	309	496	197	338	161	272	175

Notes: The unit of observation is a grade. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Female is a binary variable taking value one if a student is female. Low (high) exit grade is a high school exit grade below (above) 90 (out of 100). Robust standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

D A Difference-in-Differences Analysis of the High School Reform

This appendix applies a DiD approach to the analysis of the impact of the educational reform introduced by the Minister of Education Maria Stella Gelmini in 2010/2011, using the entire student population at Unibo.[5]

Table D.1 presents the results using a parsimonious specification where, in columns 1, 2, and 4, we only control for gender. The findings reveal that the reform does reduce STEM enrollment, aligning with the pre/post comparison observed in the TFE estimates. However, the negative effect of the reform is substantially weaker, shrinking to 6.2 percentage points compared with the more pronounced 16.8 percentage point reduction reported in Table 3. In contrast to the TFE results, the positive impact of the reform on on-time graduation (column 2) and grades (column 4) is absent in the DiD analysis. In fact, after controlling for STEM enrollment (respectively columns 3 and 5), the effect is reversed.[6]

The observed differences between the TFE and DiD estimates could, at first glance, raise questions about the external validity of the TFE approach, given that it is restricted to a subset of the broader student population. However, a critical limitation of the DiD estimator lies in its vulnerability to bias stemming from non-random selection into treatment. This limitation is particularly relevant in our context, where students' choice of high school track is likely influenced by a range of unobserved factors—such as family background, socioeconomic status, and parental preferences—that we cannot directly observe in the data. In contrast, the TFE approach inherently accounts for these confounding influences by exploiting within-pair variation among twins. Thus, while the DiD approach offers broader coverage of the stu-

[5]Formally, we estimate over the full sample the following equation, using OLS: $y_{it} = \beta_0 + \beta_1 LowMath_i + \beta_2 LowMath_i \times Post_t + \epsilon_{it}$, where the binary variable $Post_t$ refers to the post-reform period, the interaction $LowMath_i \times Post_t$ indicates whether a student attended a treated (i.e., a traditionally low-math) school after the reform, β_2 identifies the effect of the reform on treated students, and standard errors are clustered at the region of residence and high school graduation year level.

[6]Replicating these regressions with additional controls, such as fixed effects for high school graduation year and area of birth, yields nearly identical results—not shown here for brevity.

Table D.1: The effect of the high school math reform on student outcomes: DiD

	(1) STEM	(2) On-Time Graduation	(3) On-Time Graduation	(4) Grades	(5) Grades
Low Math × Post	-0.062***	-0.006	-0.016**	-0.009	-0.021**
	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Low Math	-0.265***	0.024***	-0.010*	0.126***	0.073***
	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)
Post	0.025*	0.072***	0.077***	0.075***	0.081***
	(0.014)	(0.012)	(0.013)	(0.012)	(0.012)
Female	-0.173***	0.095***	0.073***	0.148***	0.114***
	(0.007)	(0.004)	(0.004)	(0.007)	(0.007)
STEM			-0.127***		-0.196***
			(0.009)		(0.011)
Constant	0.561***	0.416***	0.488***	-0.175***	-0.066***
	(0.011)	(0.010)	(0.008)	(0.009)	(0.012)
R-squared	0.162	0.015	0.028	0.013	0.020
Observations	123015	92637	92637	1659054	1659054

Notes: The unit of observation is a student in columns 1-3 and a grade in columns 4-5. STEM is a binary variable taking value one if a student enrolls in a STEM degree. On-Time Graduation is a binary variable taking value one if a student graduates on time. Grades are (standardized) grades in exams that have been passed. Low Math is a binary variable taking value one if a student attended a low-math high school type. Post is a binary variable taking value one in the post-reform period. Female is a binary variable taking value one if a student is female. Robust standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

dent population, its susceptibility to selection bias limits the reliability of its estimates in this specific setting. The TFE estimates, although confined to a smaller and potentially less representative sample, provide a more rigorous and internally valid assessment of the causal relationships at play.