UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI

"M. FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"THE ECONOMIC RETURN OF CLASSICAL STUDIES"

RELATORE:

CH.MO PROF. GIORGIO BRUNELLO

LAUREANDO: DANIELE CORRIAS

MATRICOLA N. 1238391

ANNO ACCADEMICO 2023 – 2024

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section 'References'.

Firma (signature)

TABLE OF CONTENTS

1. INTRODUCTION

1.1 The classical lyceum and its decline

In Italy, as in many other European countries, upper secondary education is structured into vocational and academic pathways.

The vocational curricula provide a technical-scientific cultural base, favouring the development of skills that allow immediate entry into the job market. Conversely, the academic track provides broader knowledge and is aimed at continuing studies with the university. Academic education is classified in artistic, linguistic, classical, scientific and socio-pedagogical curricula.

The focus of our thesis is the classical lyceum, an institution deeply rooted in the history of the Italian society. Its path is aimed at the study of classical civilization and humanistic culture, with a strong emphasis on latin and ancient greek, compared to a smaller amount of hours dedicated to scientific subjects.

Despite having been considered the "elite" high school for decades, serving as a training ground for generations of ruling classes, we are currently witnessing an apparently unstoppable decline that has opened a debate on the usefulness, and even harmfulness, of classical lyceum.

The analysis of ministerial data shows a negative trend, with a slow but persistent decrease in enrolments over time; in 2008, the enrolment rate of the classical lyceum was 10.07%, and in the following years it fell up to 6%. Recent data for the 2023-24 school year (MIUR, 2023) indicate a further decline in enrolment, amounting to 5.8%.

The geographical distribution of enrollments (MIUR, 2023) shows how the classical studies maintain their attractiveness only in Central and Southern Italy; in Calabria, Lazio and Sicilia more than 9% of students have chosen the classical lyceum for the 2023-24 school year, a very high percentage compared to the data from Lombardia (3.7%), Veneto and Friuli-Venezia Giulia (both 3.5%).

Moreover, it is characterized by a high rate of feminization: female students are around twice as many as male students (MIUR, 2023); this aspect deserves attention, especially considering the rich scientific literature that investigates the causes of the gender gap in STEM disciplines.

The ongoing debate in Italy about classical studies was well represented and synthesized by a theatrical performance (Turin, November 14th, 2014), in which critics and supporters assumed the roles of the prosecution and the defence in a hypothetical trial, with the classical lyceum as the defendant.

According to its critics, the classical lyceum misleads its students: in fact, the belief that it is the best choice even for those who intend to pursue scientific studies is widespread in Italy, although the current decline suggests that this view has obviously diminished over the years.

The second accusation concerns its unfairness: it was conceived as an "elite school" and, at least initially, had the aim of reducing social mobility and favouring socially and culturally advantaged families.

Furthermore, individuals who exclusively pursue humanities studies may run the risk of acquiring a partial understanding of reality and lacking adequate tools to face it. This aspect is highlighted by the fact that Italians are deficient in terms of mathematical skills, so crucial for understanding today's reality. In fact, in light of the results of the PIAAC[1] survey, around 70% of Italian adults are unable to use and manipulate mathematical information effectively, compared to 52% of the OECD average[2] (Ichino, see Cardinale, Sinigaglia, 2016).

These results cannot certainly be attributed to the classical lyceum, but a contributing factor can be identified in the deeply rooted idea in the Italian society that humanistic knowledge is more important than quantitative one. On the other hand, some authors raise concerns about the other side of the coin, represented by an excess of technicality, with "pockets of hyper-specialization where an expert in rare diseases no longer knows how to treat a common cold" (Eco, see Cardinale, Sinigaglia, 2016).

More generally, the debate on the decline of the classical lyceum has stimulated an even deeper reflection: what is the purpose of secondary education? Should the usefulness of specialized knowledge or the effectiveness of methodological skills prevail? Among the defenders of the classical lyceum, the latter aspect is considered preeminent; according to Luciano Canfora (see Cardinale, Sinigaglia,2016), an authoritative historian and classical philologist, Gymnasium has the ability to arouse critical thinking, the most important of human qualities. Andrea Casalegno (2010) emphasizes the gratuitousness of the knowledge provided by classical studies, since "the impossibility of bending its acquisition to an immediate use educates and trains in that search for its own sake which is at the origin of every scientific achievement".

Finally, a rigorous debate on the topic cannot ignore the fact that the classical lyceum is an unicum of the Italian education system. In all advanced countries, latin is an optional subject chosen by few students (2-5%), while in Italy more than 30% of high school students study it

[1] Programme for the International Assessment of Adult Competencies.

[2] The data refers to the latest survey published by the OECD in 2013. Currently, the second OECD survey on the skills of the adult population (age 16-65) is underway, and its results will be released in 2024.

(Casalegno, 2010). This is due to the fact that latin is a fundamental subject not only in the classical lyceum but also in the scientific lyceum, which has experienced a surge in the enrollment in the past decades.

It is therefore important to question whether Italy should cultivate this uniqueness or come to terms with the European educational model, where classical studies have now become elective choices for future philologists.

1.2 Classical Lyceum: a historical overview

The Classical Lyceum has its origins in the Liceo Unico (also known as Gymnasium-Lyceum) established by the Casati Law (Royal Decree 13 November 1859, n. 3725 of the Kingdom of Sardinia). Its intended duration was 8 years, and it represented a post-elementary secondary education course aimed at furthering studies in the university context. It was characterized by a predominance of humanities and literary subjects. (D'amico, 2009)

In 1923, the Gentile Reform established the classical lyceum, which represented the elite school intended for the formation of future ruling classes. The prevalent function of the new school was identified as the purely cultural education of students, free from any utilitarian purposes of professional training. Only graduates with a classical education diploma were allowed free enrolment in any university faculty, while those who came, for example, from a scientific lyceum were precluded from faculties of law, medicine, arts[3] and philosophy (Cardinale, Sinigaglia, 2016). At the core of this approach was an aristocratic conception of culture and education, an elitist school reserved for a few chosen individuals.

With the Gentile reform the scientific lyceum was also established, which can be effectively defined as a humanities lyceum without Ancient Greek. Reserved for the formation of an executive, non-ruling class, it differed by a greater presence of scientific disciplines, which in the gentilian classical were instead relegated to only 15% of the total hours (D'Amico, 2009).

The gentilian structure of the classical lyceum survived the fall of fascism without significant changes. In the 1960s, classical education became a target of youth protests throughout Europe, as it represented a "hindrance to the process of democratization", as defined by French minister of education Edgar Faure in 1968 (Cardinale, Sinigaglia, 2016).

[3] More specifically, the faculty of "Lettere".

In Italy, the '68 movement denounced the elitist aspects of the gentilian model, identifying classical studies as a legacy of bourgeois culture, an enemy of the working class and the revolution[4]. In 1969, faced with pressure from the streets, the parliament passed the Codignola Law (11 Dec 1969, n.910) which liberalized access to universities, guaranteeing the possibility of enrolment in any university course with any five-year high school diploma (D'Amico, 2009). The Codignola Law decreed the loss of the elitist role of the classical high school, leading to a slow decline in the number of enrolled students which, as seen, still persists and does not seem to be stopping.

Starting from the 1970s, the classical lycei, like other upper secondary schools, enjoyed greater autonomy in defining their study plan, which led to a departure from the historical model designed by Gentile (D'Amico,2009). This phase of experimentation, made possible by the delegated decrees of 1974, generated some modifications to the classical studies curriculum, including:

- Strengthening of mathematical and natural sciences
- Enhancement of foreign language education (previously studied only in the first two years of high school)
- Adherence to the National Plan for Computer Science Education (PNI)
- Strengthening of art history education.

The aim of these educational experiments was to address some shortcomings, which in this case were mainly the absence of foreign language education in the last three years of lyceum and the limited number of hours dedicated to mathematical and scientific subjects. However, they were also criticized for causing "unbridled deregulation" (Cardinale, see Cardinale, Sinigaglia, 2016).

The experimentation phase survived until the implementation of the Gelmini reform which, starting from the academic year 2010-11, reorganized the school system by accentuating the differences between the classical and scientific licei and restoring to the classical lyceum the predominantly humanistic character that had distinguished the gentilian model.

The current organization of the classical lyceum reflects the framework outlined by the Gelmini reform in 2010. Table 1 illustrates the current study plan; the distinguishing subjects are latin and ancient greek. While latin is also studied in the traditional scientific lyceum, in the linguistic

[4] An emblematic example in this regard is an article written by Pietro Nenni in the "Avanti" newspaper which, referring to Latin, was titled "The language of the landlords".

lyceum and the traditional humanistic sciences lyceum, ancient Greek is exclusively provided by the classical lyceum.

Table 1: Classical Lyceum study plan

	1° biennio		2° biennio		
Anno	**I**	**II**	**III**	**IV**	**V**
LINGUA E LETTERATURA ITALIANA	4	4	4	4	4
LINGUA E CULTURA LATINA	5	5	4	4	4
LINGUA E CULTURA GRECA	4	4	3	3	3
LINGUA E CULTURA STRANIERA	3	3	3	3	3
STORIA			3	3	3
STORIA E GEOGRAFIA	3	3			
FILOSOFIA			3	3	3
MATEMATICA	3	3	2	2	2
FISICA			2	2	2
SCIENZE NATURALI	2	2	2	2	2
STORIA DELL'ARTE			2	2	2
SCIENZE MOTORIE E SPORTIVE	2	2	2	2	2
RELIGIONE CATTOLICA **ATTIVITA' ALTERNATIVE**	1	1	1	1	1
TOTALE ORE SETTIMANALI	**27**	**27**	**31**	**31**	**31**

For the purpose of this research project, it is useful to compare the curriculum of the classical lyceum with that of the scientific lyceum. The most significant differences concern:

- Latin, which is provided for 3 hours per week in the scientific lyceum.
- Mathematics, to which the scientific lyceum devotes 5 weekly hours in the first two years and 4 in the last three years.
- Physics and natural sciences, both studied for 2 hours per week in the first two years and 3 in the last three years.
- Ancient Greek, not taught in the scientific lyceum.

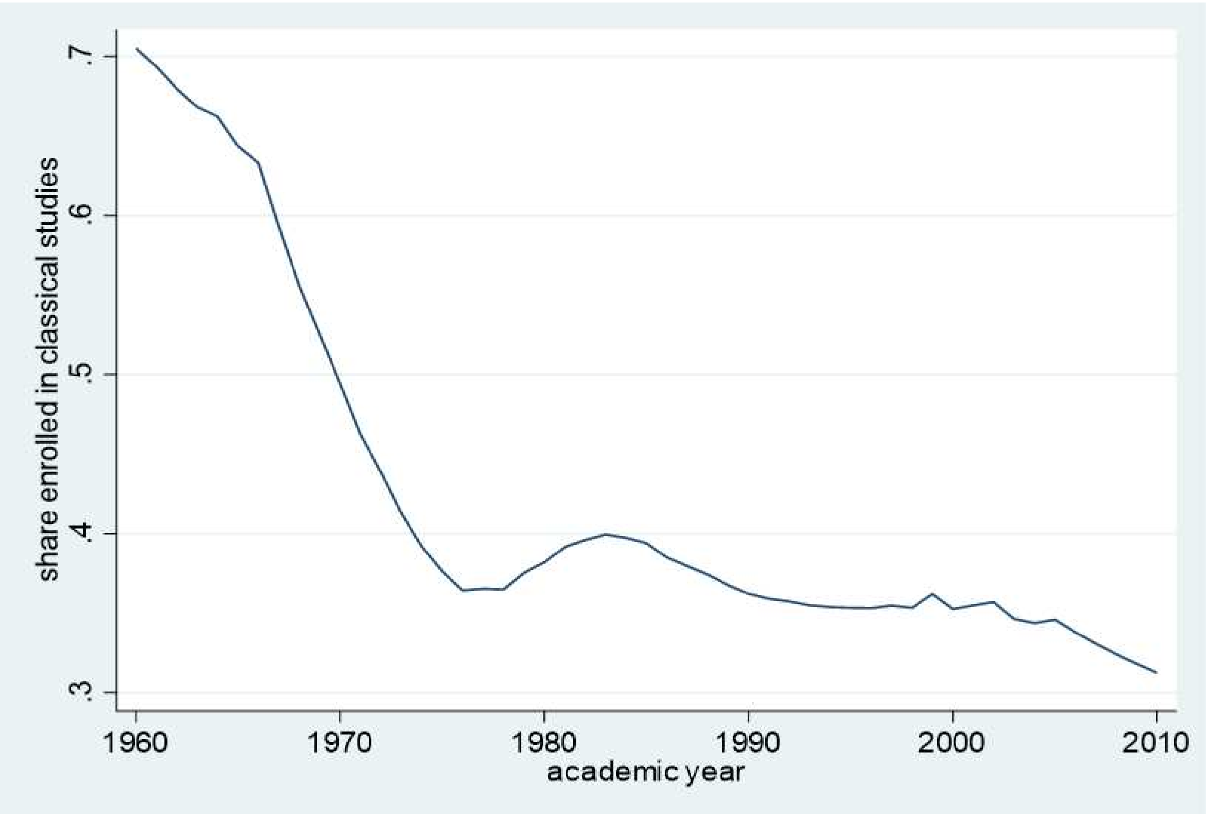
Such differences would become even more pronounced if we were to compare the classical curriculum to the new directions of the scientific lyceum, such as the Applied Sciences

Lyceum[5], which in recent years has seen a significant increase in enrollments, even at the expense of the traditional scientific lyceum. The shift of students towards this type of high school, in which the teaching of Latin language and literature is not even included, seems to demonstrate how, in their perception, the study of classical languages requires too much time and energy to acquire skills that are not easily transferable in the future, in a job market that is more focused on mathematical and scientific competencies.

The decline of classical Lyceum becomes even more evident if, instead of observing historical trends relative to the total number of students enrolled in upper secondary education, we focus on those who opt for an academic education.

Within the academic track, the share of students enrolled in the classical lyceum has collapsed from more than 70% in 1960 to close to 30% at the beginning of the current century, as shown in Figure 1 (MIUR, see Brunello, Rocco et al., 2023).

Figure 1: share of students enrolled in classical studies within academic track



[5] In italian language known as Liceo scientifico – opzione scienze applicate.

Chapter 2- LITERATURE REVIEW

The effect of education on economic variables has been extensively investigated in the literature over the past decades. However, there are currently no previous studies regarding the economic impact of attending a classical lyceum. What reasons can explain this absence?

First and foremost, the classical lyceum is a distinctly Italian institution, without counterparts in other educational systems where classical studies are relegated, as previously mentioned, to optional supplementary courses. Additionally, in Italy there is a scarcity of datasets containing detailed information on pre-university educational paths, particularly regarding the high-school track. This issue is shared with other European countries; for instance, the Labour Force Survey, one of the richest and most used datasets, collects information only on the highest level of education attained[6].

A second reason is that the majority of the existing literature on the economic consequences of high school choice primarily focuses on the comparison between academic and vocational education (see Brunello, Rocco 2017, Hanushek 2017, Ollikainen 2022), rather than within different types of academic education.

The only study on the classical lyceum that we can mention is the one conducted by Brunello, Rocco et al. (2023) titled “Do Classical Studies open your mind?”. Unlike the present work, the paper investigates the effect of attending classical studies not on economic and labour market variables, but rather on the so-called Big Five, personality traits identified by the psychological theory: conscientiousness, openness, agreeableness, extraversion, and neuroticism.

In their paper, as in the present thesis work, data from PLUS (Partecipation, Labour, Unemployment Survey) were used. However, they focused exclusively on the 2018 wave, which was the only one containing a module on non-cognitive skills, used to construct the aforementioned Big Five Personality Traits.

The final sample is composed of individuals who have completed either a classical or a scientific lyceum; the two groups are not identical but similar, since they attract student drown from the upper part of both the parental background and the pre-enrolment ability distributions.

[6] In the Labour Force Survey dataset, if the highest level of education achieved by a student is the college degree, there are no informations on his high school diploma.

Using entropy balancing approach and controlling for a set of observable characteristics, they find interesting results which are reported on Table 2.

Table 2: The effect of a classical curriculum on outcomes (Brunello, Rocco et al. , 2023)

Variables	Completed College	Probability of employment	Log (1 + annual earnings)	Unhappy	Openness	Agreeableness	Conscientiousness	Extraversion	Neuroticism
Panel A. *Entropy Balancing*									
	0.0337***	0.0000	-0.0358	0.0274***	0.0543	-0.0308	0.0227	0.0844	0.2350***
Classical curriculum	(0.0099)	(0.0110)	(0.1106)	(0.0085)	(0.0590)	(0.0545)	(0.0545)	(0.0717)	(0.0665)
Oster beta	0.0336	0	-0.0357	0.0274	0.0542	-0.0308	0.0227	0.0844	0.235
Westfall test	.015								
R-squared	.0678	0.1302	0.1405	0.0246	0.0277	0.0325	0.0296	0.0264	0.0303
Panel B. *Propensity score matching*									
	0.0297***	-0.0044	-0.0740	0.0270***	0.0369	0.0271	0.0638	0.0978	0.2478***
Classical curriculum	(0.0106)	(0.0114)	(0.1148)	(0.0085)	(0.0622)	(0.0562)	(0.0552)	(0.0724)	(0.0680)
Mean	.746	0.663	6.621	0.134	9.398	10.75	11.65	7.729	6.072
Observations	8,299	8,299	8,299	8,299	8,299	8,299	8,299	8,299	8,299

Students from classical lyceum (with respect to the control group: students with a scientific high-school diploma) are more neurotic[7] (+3,87%) and are 20.4% more likely to report unhappiness, while no statistically significant effects have been found on conscientiousness and openness.

The author's conclusion contrasts with the idea that classical studies increase the mental openness, which is one of the main arguments used in defence of the classical lyceum.

Moreover, the paper also investigated the effect of high school choice on educational attainments and some economic outcomes. As shown in Table 2, students with classical studies are 4.52% more likely to graduate from college, even if this is not reflected in better labour market outcomes; the authors consider this latter aspect as one of the possible explanations for the negative effect that graduating from a classical lyceum has on self-perceived happiness, as the expectations generated by the college degree are not met with an adequate economic return.

Although there is no "direct" literature on classical studies, there exists a vast body of research that has examined the effect of high school curricula on future economic outcomes. Particularly, in a labour market where scientific skills are increasingly required, many studies have focused

[7] Trait disposition to experience a set of negative moods such as anger, irritability, anxiety, depression (Brunello, Rocco et al.,2023).

on the economic impact of incorporating more hours of math-oriented disciplines during the upper secondary education.

Since the difference between classical and scientific lyceum can be seen in terms of variations in the amount of time dedicated to mathematical and scientific subjects, we consider important to delve into this branch of the economic literature. In other words, if we were to find that attending the scientific lyceum (as opposed to a classical lyceum) has a positive and statistically significant effect on the economic outcomes under consideration, we could interpret our results as follows: "increasing the hours dedicated to mathematics, physics and sciences at the expense of latin and greek has a causally positive economic effect".

For many years, the lack of data and information has hindered investigations into the link between high school curricula and wages. One of the earliest findings in this research field was accomplished by Altonji (1992), who used the National Longitudinal Survey of the High School Class of 1972 (NLS72), an American dataset which contains information on students' semester-based course load during grades 10-12 and their wages up to 13 years later.

He regresses the log-wage of each person on a vector of credits completed in eight different subjects [8]during grades 10-12, along with a set of covariates. The main issue faced by the author has been the presence of unobserved differences in pre-high school ability and preferences for post-secondary education; as shown in the paper, ability and achievement measures positively correlate with semester hours of math, science, foreign language and are negatively correlated with semester hours of commercial arts and industrial arts. Since abler pupils self-select themselves into math, science and foreign language courses, Ordinary Least Squares (OLS) estimates of the high school curriculum[9] effect on wages would be, presumably, upward biased.

In order to solve the omitted variable bias, Altonji (1992) exploits the variation in curricula across 897 US high schools using the school's mean number of credits earned by students within each subject as an instrument for each student's number of credits earned in that subject.

Robust across estimation methods (OLS, OLS with school fixed effect, IV), he concludes that the return to a year of high school courses is smaller than the value of one year spent in high school, and therefore the curriculum effect is "extremely weak".

The Table 3 below illustrates the effects of high school curriculum on wages. Numerically, an additional year composed only of Math, Science and Foreign language classes (leaving aside

[8] Science, Mathematics, Foreign Language, Social studies, English, Industrial Arts, Commercial Arts, Fine Arts.
[9] Building a curriculum consisting of Mathematics, Science and Foreign Language.

subjects such as English, Social studies and vocational courses, for which the estimates observe a negative effect on wages) increases wages by 3.1 percent. In a different specification (Table 3, column 2) where Altonji controls for family background, the effect of course combination falls to 1,7%. In any case, those numbers are far below the marginal return to education, around 7%.

Table 3: The effects of High School Curriculum on Wages (Altonji,1992) – Dependent variable: Log Average Hourly Wage in 1967 Dollars

	Combined Sample1								
	Instrumental Variables Estimates				OLS			OL, BS Fixed Effects	
Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Effects of Course Combinations									
10 hrs. each of Math, Science Foreign Language	.031 (.013)	.017 (.012)	.019 (.012)	.013 (.013)	.035 (.006)	.033 (.006)	.016 (.007)	.044 (.008)	.019 (.009)
10 hrs. each of the 5 academic subjects	.003 (.014)	-.009 (.013)	-.005 (.014)	-.012 (.014)	.004 (.009)	.003 (.009)	-.007 (.009)	.025 (.013)	.005 (.013)
Avg. number of courses/year in the 5 academic subjects	-.016 (.011)	-.020 (.010)	-.018 (.011)	-.020 (.011)	-.009 (.007)	-.010 (.008)	-.013 (.007)	.008 (.011)	-.003 (.011)
Avg. number of courses/year in the 8 subjects	-.013 (.011)	-.016 (.010)	-.014 (.010)	-.016 (.010)	-.009 (.008)	-.010 (.008)	-.012 (.008)	.009 (.011)	-.002 (.011)
Controls2									
Controls for region, city size College proximity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes3	Yes[3]
Controls for family background4 Parental SES variables	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes
Parental involvement and aspirations	No	Yes	Yes	Yes	No	Yes	Yes	No	No
Father, mother want college	No	No	No	No	No	No	No	No	No
Controls for aptitude and achievment[4]	No	No	Yes	No	No	No	Yes	No	Yes

1 Corrected standard errors are in parentheses. They account for arbitrary High School specific patterns of correlation and heteroskedasticity. The estimators are defined in the text.
2 For a list of the control variables see the text and Appendix Table A.1.
3 Equation includes high school specific intercepts, which implicity control for all high school specific variables, including location.
4 In the IV case both values for the individual and high school means are included. In the OLS case only individual values are included.

In this paper, titled "The effect of high school curriculum on education and labour market outcomes", Altonji(1992) anticipated potential external critics by thoroughly describing all the limitations of the econometric approach chosen. In particular, the instrument is likely to be endogenous since it probably correlates with some variables as primary school preparation, average family background, quality of the courses and average student's ability, which are all imperfectly controlled in the model.

Among the goals set by the author was also an attempt to answer the question "Do years spent in school have economic value?", by delving into the age-old debate between the Human Capital interpretation and the Signaling Theory. According to the Human Capital Theory, education allows individuals to acquire skills that make them more productive in the labor market, consequently leading to higher earnings. In contrast, according to the Signaling model, more capable individuals attend advanced courses and attain a higher level of education (both of which are unavailable to less able individuals) not because this will result in greater productivity, but to provide a signal to the labor market.

The results of the paper, if taken at face value, seem to loan toward the latter aspect: higher education does not have a significant impact on human capital but serves as a signal to the labour market.

In 1995 Levine and Zimmermann made their contribution by estimating the effect of high school math and science courses on wages. They used not only the NLSY, the source of data already exploited by Altonji (1992), but also the senior cohort of the High School and Beyond Transcript Dataset (HSB), most of whose members graduated in 1980.

Unlike the NLSY, the HSB only follows individual's income up to six years after graduation; this negative aspect should be emphasized because an individual may not have fully settled into careers within this time frame. Moreover, it is difficult to conceive that income at six years post-graduation could in any way substitute the life-time income, which, if available, would be the ideal economic outcome in our research questions.

They estimated by Ordinary Least Squares (OLS) the Equation 1 displayed below, where the main outcome of the model, represented by the subsequent labour market earnings *ln Wi* , is regressed on the number of credits earned in math and science courses Xi, a set of family background characteristics Fi, a set of observed individual characteristics Ci, along with a test score measure, Si, used to proxy the unobserved pre-enrollment individual's ability .

$$\ln Wi = Xi\beta_1 + Fi\,\beta_2 + Ci\beta_3 + Si\beta_4 + ei \qquad (1)$$

Tables displayed below shows the findings coming from the NLSY source of data (Table 4) and the HSB dataset (Table 5). The results provide initial alternative evidence to the null effect shown by the pioneering work of Altonji (1992): in both datasets and consistent across different specifications, Levine and Zimmermann found that the math effect was positive and statistically significant but only for a subgroup of the population, consisting of highly educated women. In particular, proxying the unobserved ability through a vector of test-score measures, an additional high school semester of math increases log wages for female graduates by 5.4% (see column 7, Table 5); using instead the NLSY dataset, the magnitude collapses, remaining statistically significant (Table 4).

Table 4: The effects of the High School Math and Science Curriculum on Log Weekly Wages of full-year workers, by gender and educational attainment: NLSY Data (Levine and Zimmermann, 1995)

Variables	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
Curriculum variables	**All workers[b]**					
%Math	.709 (.324)	.498 (.332)	.280 (.344)	1.075 (.390)	.679 (.381)	.485 (.383)
%Science	.344 (.345)	-.216 (.350)	-.055 (.355)	.285 (.394)	.021 (.394)	.056 (.395)
Sample size	1,891	1,822	1,822	1,453	1,427	1,427
	High-school graduates					
%Math	.596 (.441)	.654 (.463)	.500 (.466)	1.176 (.713)	.814 (.686)	.703 (.672)
%Science	.412 (.479)	.041 (.489)	.188 (.507)	-.136 (.704)	-.216 (.702)	-.059 (.704)
Sample size	953	911	911	617	602	602
	College graduates					
%Math	.758 (.588)	.304 (.625)	-.324 (.608)	2.258 (.597)	1.778 (.583)	1.431 (.619)
%Science	.667 (.603)	.327 (.626)	-.020 (.648)	.306 (.527)	-.103 (.500)	-.012 (.490)
Sample size	506	494	494	418	415	415
Additional control variables						
AFQT score[c]	No	Yes	No	No	Yes	No
Individual test scores[c]	No	No	Yes	No	No	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: The effects of the High School Math and Science Curriculum on Log Weekly Wages of full-year workers, by gender and educational attainment: HSB Data (Levine and Zimmermann, 1995)

Variables	Men				Women			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Curriculum variables								
				All workers				
#Math	.028 (.011)	.030 (.012)	.028 (.012)	-.017 (.063)	.023 (.008)	.023 (.009)	.019 (.009)	-.060 (.051)
#Science	.009 (.009)	.014 (.010)	.013 (.010)	-.029 (.058)	.002 (.009)	.002 (.010)	.001 (.010)	.066 (.046)
Sample size	2,008	1,828	1,801	1,801	2,325	2,150	2,108	2,108
				High School graduates				
#Math	.028 (.013)	.031 (.014)	.031 (.014)	-.0781 (.082)	.008 (.009)	.004 (.010)	.003 (.010)	-.038 (.070
#Science	-.002 (.011)	.005 (.012)	.005 (.012)	-.041 (.078)	.004 (.010)	.003 (.010)	.003 (.010)	.031 (.067)
Sample size	1,177	1,068	1,054	1,054	1,255	1,150	1,131	1,131
				College graduates				
#Math	.017 (.022)	.037 (.022)	.029 (.023)	-.030 (.109)	.056 (.021)	.056 (.022)	.054 (.024)	-.184 (.115)
#Science	.033 (.020)	.020 (.019)	.017 (.020)	-.087 (.103)	.012 (.017)	.013 (.018)	.008 (.018)	.030 (.071)
Sample size	453	419	415	415	574	541	529	543
Additional control variables								
Composite test scores	No	Yes	No	No	No	Yes	No	No
Individual test scores	No	No	Yes	Yes	No	No	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation technique	OLS	OLS	OLS	2SLS[c]	OLS	OLS	OLS	2SLS[c]

For what concerns the other subgroups, the results are not consistent across different datasets; in the HSB sample (Table 5), an additional semestral math course increases wages by 3.1%, but the return becomes not statistically significant using the NLSY sample (Table 4). Some authors (Rose, 2004) observed, however, that the variation in results across subgroups may be due to the small sample size within each group.

On the other hand, additional courses in science do not increase subsequent earnings in a statistically significant way, and this conclusion holds for any subgroup in both the datasets.

Furthermore, Levine and Zimmermann replicate the Instrumental Variables approach (IV) used by Altonji (1992): using as an instrument the mean number of math and science classes taken by students in each high school, also the math effect disappears.

Another contribution to the literature on this subject was made in 2004 by Rose and Betts, in their paper titled "The Effect of High School Courses on Earnings". This work differs from previous studies due to the presence of highly detailed transcript data, containing information about the types of math courses taken by students. In particular, the authors identify eight different math courses: vocational math, pre-algebra, algebra/geometry, intermediate algebra, advanced algebra, calculus.

As in the paper by Levine and Zimmerman (1995), also in this case the data source is represented by the High School and Beyond (HSB) dataset. However, Rose and Betts (2004) exploit information from the cohort of second-year high school students. This choice offers several advantages:

- The possibility to include some high school dropouts in the sample, who were excluded from the dataset used by Levine and Zimmerman (1995), which focused on the senior cohort.
- The ability to track individual's earnings up to ten years after high school graduation because the majority of the sample graduated in 1982. This allows for a rich presence of earnings data for those students who pursued higher levels of education.

The authors construct a log-linear model (see Equation 2) in which the log of 1991 annual earnings, for each individual i, is regressed on a vector of credits earned in the six types of math courses. The objective is to capture the causal effect, interpreted as a percentage increase in earnings due to a marginal increase in the number of credits in a specific math course. To achieve this, Rose and Betts (2004) use a rich set of control variables, including individual, family and school-related characteristics; they also include a set of dummies that capture the highest level of education attained by each individual in 1992.

$$\ln earn_{is} = \alpha + ß_0 Curric_{is} + ß_1 Demo_{is} + ß_2 Fam_{is} + ß_3 Sch_{is} + ß_4 HiDeg_{is} + \varepsilon_{is} \qquad (2)$$

In a different specification, the authors attempted to address the bias stemming from the difficulty to measure student's cognitive ability (a variable that, as we will see later, is correlated both with educational choices and economic success) and their motivation. They used the student's GPA (math grade point average) in math courses as an approximation for these factors.

When controlling for demographic, family and school-related characteristics, as well as the highest level of education achieved, their findings indicate that a unitary increase in credits earned in an algebra/geometry course is associated with an estimated 3.1% increase in earnings. This effect becomes more pronounced (Table 6) with a calculus course (6.5%) and an advanced algebra course (4.2%).

Table 6: The effects of specific Math Courses on Log-Earnings: OLS, IV and Fixed Effects Estimates (Rose, Betts, 2004)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Vocational	0.001 (0.011)	-0.024** (0.011)	-0.027** (0.010)	-0.029** (0.011)	-0.084** (0.030)	-0.086** (0.030)	-0.019 (0.012)	-0.023* (0.013)
Pre-Algebra	0.067** (0.014)	0.023 (0.014)	0.007 (0.014)	0.005 (0.014)	0.015 (0.034)	0.013 (0.034)	0.006 (0.017)	0.004 (0.017)
Algebra/Geometry	0.080** (0.010)	0.061** (0.010)	0.031** (0.010)	0.029** (0.010)	0.090** (0.035)	0.083** (0.035)	0.029** (0.012)	0.027** (0.012)
Intermediate Algebra	0.109** (0.017)	0.078** (0.016)	0.032** (0.016)	0.022 (0.017)	-0.107 (0.068)	-0.100 (0.67)	0.054** (0.019)	0.042** (0.019)
Advanced Algebra	0.134** (0.014)	0.088** (0.014)	0.042** (0.014)	0.029** (0.015)	-0.077 (0.050)	-0.082 (0.050)	0.054** (0.017)	0.039** (0.017)
Calculus	0.195** (0.033)	0.120** (0.032)	0.065** (0.032)	0.047 (0.032)	-0.132 (0.167)	-0.140 (0.167)	0.077** (0.036)	0.058 (0.036)
Total Math Credits	0.106** (0.006)	0.069** (0.007)	0.027** (0.007)	0.019** (0.007)	-0.009 (0.024)	-0.010 (0.024)	0.036** (0.008)	0.027** (0.008)
Math GPA				0.036** (0.009)		0.063** (0.015)		0.039** (0.009)
Other Controls								
Demographic Information	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	Yes	Yes	Yes	Yes	No	No
Highest Educational Degree	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Estimation Method	OLS	OLS	OLS	OLS	IV	IV	FE	FE
R-squared	0.069	0.172	0.199	0.201	0.187	0.192	0.316	0.319
Number of Obs	5.919	5.919	5.919	5.896	5.864	5.841	5.919	5.896

In Column 4 of Table 6 the authors illustrate that, when controlling for the effects of cognitive ability, the coefficients decrease in magnitude, and the returns for some courses (intermediate algebra and calculus) are no longer statistically significant. However, OLS results indicate a marginal return of 2.9% for both algebra/geometry and advanced algebra.

Furthermore, Rose and Betts (2004) attempt to construct an instrument similar to the one used by Altonji (1992) to compare the two sets of results, finding that the marginal return for an algebra/geometry course is estimated to be 8.3%.

In conclusion, their findings are consistent with various econometric approaches and model specifications:

- Math courses have a causal effect on future labor market earnings

- More advanced courses appear to have a greater impact than less-academic courses.

Although Rose and Bett's paper (2004) represents a significant improvement over previous studies in this research field, it fails to address some recurring issues:

- OLS estimation results are unable to identify causality in the relationship between educational choices and the dependent variables under examination, due to the presence of relevant unobservable factors, such as pre-enrolment ability.
- The instrumental variable designed by Altonji (1992) and also used in other papers seen so far (Levine and Zimmerman, 1995, Rose and Betts, 2004) does not represent a clean natural experiment, because it is correlated with several variables that are not perfectly controlled in the models and that influence both educational choices and future economic outcomes.

An important contribution to the debate comes in 2006 from Joensen in a paper titled "Is there a causal effect of high school math on labor market outcomes?". He uses the exogenous variation induced by the Danish high school pilot scheme, an experimental curriculum applied to a certain number of high schools in Denmark before 1988.

The sample examined consists of the 1986-87 high school Danish cohorts. Before the 1988 structural reform, students commenced their upper secondary education by choosing between the Math track and the Language track. At the end of the first year, students from the Language track could choose one of the following four branches: SocSci-Language, Music-Languages, Modern languages, Classical languages. Simultaneously, students from the Math track could choose between Math-SocSci, Math-NatSci, Math-Music, and Math-Physics (Joensen, 2006).

Focusing on students in the math track, it is important to highlight that, upon high school graduation in Denmark, each student was assigned a rating for their level of mathematics achievement, distinguishing between high, medium and low levels. However, the only way to attain an advanced level of mathematics was by choosing the Math-Physics package.

The pilot scheme, implemented only in select schools, provided math track students with an additional option to attain a high level of math: the Math-Chemistry package. In other words, the pilot scheme represents an exogenous variation induced by a policy, which reduces the cost of attaining an advanced level of mathematics. This mechanism occurs because the possibility of combining advanced math with chemistry is of interest to a portion of students, those who would choose advanced math courses if they were not deterred by the difficulty of physics, as also noted by other economists (Albaek, 2003, see Joensen, 2006). This exogenous effect, which simulates a natural experiment by making the curriculum choice as good as random, is used by Joensen as an instrumental variable. The instrumental variable approach (IV), under certain conditions, allows the identification of the Local Average Treatment Effect (LATE), which represents causal effect of high-level math on earnings (in log) for the subgroup of individuals who are induced to choose advanced math because they were exposed to the pilot scheme and who, otherwise, would not have chosen that level of math.

The first section of Table 7, depicted below, displays the results obtained using a dummy variable named *PilotSchool,* which equals 1 if the individual is assigned to the pilot scheme and 0 otherwise. In the primary model specification, the natural logarithm of labor market income thirteen years after starting high school is regressed on a treatment variable known as *MathA,* which equals 1 if the individual chooses the high-level math course and 0 otherwise, along with a set of control variables.

When controlling for individual's ability (measured through their grade point average, GPA), parental background and demographic characteristics, the causal effects of advanced math courses on earnings appear to be positive and statistically significant. In particular, students who opt for the high-level math courses influenced by the pilot scheme earn an income that is 32% higher than others. However, as indicated in Column 5 of Table 7, part of this effect is indirect and arises from the fact that pursuing advanced math also impacts the level of education attained.

Table 7: Estimation of the Causal Effect of High Level Math on Labor Market Income for High School Graduates (Joensen, 2006)

Variables	Parameter estimates, (standard errors) and (Marginal effects)													
	IV Pilot School							IV Dist Pilot School (10km)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(06)**	**(07)**	**(1')**	**(2')**	**(3')**	**(4')**	**(5')**	**(6')**	**(7')**
Outcome Equation:														
High Level Math	0.26 (0.08)	0.29 (0.07)	0.32 (0.06)	0.12 (0.06)	0.13 (0.06)	0.20 (0.06)	0.21 (0.06)	0.32 (0.22)	0.47 (0.14)	0.51 (0.13)	0.19 (0.12)	0.21 (0.12)	0.47 (0.11)	0.47 (0.11)
First Stage:														
IV (Pilot School or Dist Pilot School	0.41 (0.02) [0.14]	0.44 (0.02) [0.14]	0.45 (0.02) [0.14]	0.44 (0.02) [0.14]	0.44 (0.02) [0.14]	0.48 (0.02) [0.15]	0.48 (0.02) [0.15]	-0.11 (0.01) [-0.03]	-0.11 (0.01) [-0.03]	-0.11 (0.01) [-0.03]	-0.11 (0.01) [-0.03]	-0.11 (0.01) [-0.03]	-0.12 (0.01) [-0.04]	-0.12 (0.01) [-0.04]
Additional Control Variables:														
GPA			+		+		+			+		+		+
Individual Variables:														
Gender		+	+	+	+	+	+		+	+	+	+	+	+
Experience (quadratic)		+	+	+	+	+	+		+	+	+	+	+	+
Educational Variables:														
Educational Level grouping				+	+						+	+		
Educational Level-subject grouping						+	+						+	+
Parental variables (for mother and father): Highest completed education , and income		+	+	+	+	+	+		+	+	+	+	+	+

In addition, the author constructs another instrument named *DistPilotSchool,* which is a continuous variable measuring the distance between a high school implementing the pilot scheme and the nearest "standard" high school. As demonstrated in Table 7, the coefficient's magnitude is substantial, but their precision is quite low.

Nevertheless, Joensen (2006) makes a significant contribution to the debate on high school curricula: math courses exert a strong influence on future labour market outcomes. However, certain issues persist, as highlighted by the author:

- The findings of the paper cannot be deemed applicable to the entire population, as the Instrumental Variable approach does not allow for the identification of the Average Treatment Effect (ATE).
- Schools are not randomly selected for participation in the pilot scheme.
- The dependent variable is the natural logarithm of earnings thirteen years after enrolling in high school, whereas the preferred income measure would be lifetime earnings.

As we will see in Chapter 3, some of the issues encountered will recur in the empirical work that forms the core of this thesis.

3. EMPIRICAL ANALYSIS

3.1 The Dataset

In order to investigate our research question, we leverage the information contained within the PLUS survey, produced by INAPP (National Institute for the Analysis of Public Policies), an Italian national research institute.

PLUS (Participation, Labour, Unemployment Survey) is a recurring national sample survey of over 50.000 individuals aged 18-74. It was initiated in 2005; however, in this paper we focus on the most recent three waves: 2014, 2016, 2018.

The PLUS survey primarily focuses on the analysis of the national labour market. As described on the official INAPP website, the research investigates several specific aspects, including the entry of young individuals into the labour market, job search methods, and the participation of women in the workforce.

Why have we chosen this survey?

As previously mentioned, the primary data sources used in micro-econometric analysis (such as the quarterly Labour Force Survey) lack sufficient information about each individual's educational trajectory. In contrast, the PLUS survey not only encompasses data on individuals' highest educational attainment but also includes information on their educational pathways. Specifically, it allows us to identify the type of high school chosen and attended by each individual and consequently construct our key treatment variable, distinguishing those who pursued classical studies from those who completed scientific studies.

Moreover, this dataset also includes data on each individual's lower secondary education exit score, which, as seen in the literature review, can be used to approximate the pre-enrollment unobserved ability, a primary factor affecting educational choices. Lastly, it enables the use of information about the socio-economic context in which the individual was born and raised.

The three considered questionnaires are not entirely identical, because certain questions of interest were included only in specific waves and not in others. For instance, information regarding the junior high school exit score were collected in 2014 and 2016, but not in 2018.

The initial sample consisted of 114.240 observations: 47.507 coming from the 2014 questionnaire, 30.788 from 2016 and 35.945 from 2018.

The first issue addressed during the dataset cleaning process pertains to the presence of a panel component: some individuals were interviewed multiple times in different years. In such cases, we retain only the most recent observation and eliminate the panel component from the sample.

The final sample comprises individuals who possess either a classical or a scientific high school diploma, are no longer full-time students, were born between 1951 and 1995 and were raised in Italy.

Our final working sample consists of 15.298 observations, divided across the years 2014, 2016 and 2018 as illustrated below in Table 8.

Table 8: Final working sample by year

year	Freq.	Percent	Cum.
2014	3,803	24,86	24,86
2016	6,846	44,75	69,61
2018	4,649	30,39	100,00
Total	15,298	100,00	

3.2. The Selection bias

Our main goal is to compute the return to high-school in terms of educational attainments, future labour market outcomes and other non-monetary life-style outcomes. Comparing the returns associated to classical studies and those associated to scientific studies, it is possible to see if a gap exists, what is its size and if it is relevant from a statistical point of view.

Suppose that Y (C) is the outcome related to classical studies and Y (S) the outcome related to scientific studies. For each individual we can observe only one of the two potential outcomes, and therefore it is impossible to identify the individual gap between the return of classical lyceum and scientific lyceum, which would be [Y (C) – Y(S)].

Given that it is not possible to ascertain the counterfactual for any individual, the second step involves considering the following: what would be the ideal experiment capable of capturing the causal effect in the reference sample (Angrist, Pischke, 2008)?

Ideally, by taking all junior high school's graduates and randomly assigning them to either a classical lyceum or a scientific one, in the presence of large numbers, we would obtain two groups in such a way that one would serve as the counterfactual to the other. Under random assignment, the two groups would be homogeneous with respect to all relevant[10] characteristics, with a small sampling error dependent on the sample size (Angrist, Pischke, 2008).

However, a typical issue that economists address is that in this field we usually work with observational data and not with experimental data: this implies that a simple comparison of labour market outcomes between classical and scientific students is not able to capture the causal effect of high-school choice because of selection bias. Individuals are not randomly assigned to scientific or classical studies: they self-select themselves into high school based upon personal and familiar characteristics.

Therefore, our fundamental question is: how similar are the treatment group (students with a classical diploma) and the control group (students with a scientific diploma)? We try to answer this question based on the observable characteristics.

[10] "Relevant characteristics" means that they influence the dependent variables and are correlated with the treatment variable.

3.3 Descriptive Statistics

What are the main characteristics that affect this enrolment choice?

A recurring key factor in the previous studies is undoubtedly represented by the socio-economic background: the social, cultural and economic context of origin strongly affects the choice of the high-school type and, more in general, the educational pathways of each individual. In addition, parental background also has a direct impact on the income an individual will earn during its career, for example, by providing a network of connections or allowing students to focus on their university studies without the need to work to support themselves (Brunello, Rocco, 2021).

Table 9 presents the distribution of maternal education by type of high school degree.

In this section dedicated to descriptive statistics, we choose to enrich the sample by including also individuals who hold high school diplomas from neither the classical lyceum nor the scientific lyceum. Specifically, all high school diploma holders who do not have a background in the classical or scientific lyceum are categorized together in a third group referred to as "Other".

The "Other" group primarily consists of students who have pursued vocational education. The rationale for this choice becomes evident upon looking at Table 9: the most significant disparities in terms of parental education revolve around the differentiation between vocational and academic education[11]. Individuals with classical and scientific diploma are clearly drawn from the upper part of the maternal education distribution.

The data demonstrates a strong correlation between parental education and the high school track: less than 3% of individuals belonging to the third group, referred to as "Other", have mothers with a college degree, and furthermore, only in 19,54% of cases do mothers possess a high school diploma.

Looking within academic studies, are there relevant differences between classical and scientific students in terms of maternal education? Over 22% of students with classical studies have a mother who graduated from college; this percentage falls to 15,51% for individuals with scientific studies.

These data align with what was described in the introduction of this paper: students from classical and scientific tracks present a more privileged family background compared to their peers with vocational educational (mostly from "istituti tecnici e professionali").

However, as a portion of our sample made the high school choice during a historical period when the classical lyceum was considered the elite institution (Cardinale, Sinigaglia, 2016), it is not surprisingly that classical students exhibit a better family background even compared to those in the scientific track.

[11] Both classical lyceum and scientific lyceum are categorized as academic education.

Table 9: Distribution of maternal education by type of high school degree

Maternal highest degree	(1) classic	(2) scientific	(3) other
Missing	2,55	2,76	4,04
Elementare	16,27	18,08	42,36
Media inf.	20,02	24,52	31,23
Diploma	38,46	39,13	19,54
Laurea	22,70	15,51	2,78
Total	100	100	100

The distribution of paternal education by type of high school degree presents similar results, as illustrated in Table 10: in this case as well, while the elite background of both classical and scientific students is evident (if compared to the remaining part of the student population), it is noteworthy that 30,43% of classicists have a father with a college degree, compared to 20,04% of those who attended the scientific lyceum.

Table 10: Distribution of paternal education by type of high school degree

Paternal education	(1) classic	(2) scientific	(3) other
Missing	2,99	3.16	5.05
Elementare	13.08	14.32	36.71
Media inf.	19.28	24,52	32.76
Diploma	34.23	37.96	21.58
Laurea	30.43	20.04	3.90
Total	100	100	100

Data on parental occupation confirms our analysis : as shown in Table 11, fathers of those who attended the classical lyceum have a high-ranked occupation (managers, professionals, white collars) in 66.56% of cases, a share that decreases to 60,83% for students in the scientific track.

Table 11: Distribution of paternal occupation by type of high school degree

Paternal occupation	(1) classic	(2) scientific	(3) other
Managers	26.89	20.10	8.99
Professionals	11.11	9,75	4.72
White collar	28.56	30.98	19.78
Services and trade	11.30	13.03	16.93
Blue collar	14.76	18.42	38.65
Elementary	1.95	2.03	4.92
No job or military	2.97	3.10	2.66
Don't know	2.46	2.60	3.36
Total	100	100	100

These differences, albeit in the realm of correlation and not causality, suggest that we are comparing two similar but not identical groups. Therefore, it is essential to use a control variable that captures the socio-economic background from which each individual comes.

Building upon the work presented in the paper "The Pathways to College" (Brunello, Rocco, 2021), we construct an index of parental background, called *famback,* by applying principal component analysis to the following variables:

- Mother's and father's years of completed education
- Mother's and father's occupation (we derive the variables based on the ISEI-08 Index of occupational status)
- Minimum parental occupation and educational level.

Principal Component Analysis (PCA) is a multivariate statistical technique used to reduce the dimensionality and complexity of a set of related data while preserving the maximum variance present in the data. In other words, PCA aims to transform the original variables into a new set of uncorrelated variables called "principal components", which capture the main sources of variation in the data (Stock, Watson, 2007).

Further details on the construction of the variable *famback* through Principal Component Analysis are provided in the Appendix.

Another key aspect affecting this choice is the pre-enrollment individual's cognitive ability, a typically unobservable variable. The chosen solution is to use the junior high school exit grade as a proxy for the unobserved cognitive ability.

Table 12, reported below, illustrates that in the Italian school system students with the highest cognitive ability (which, ideally, are reflected in junior high school performances) tend to choose either the classical or the scientific track. Therefore, they are clearly drawn from the upper portion of the pre-enrollment ability distribution.

In our sample, 59,19% of classical students completed the lower secondary education with the highest grades, compared to 52,69% among scientific students. Among those who choose other schools (mostly vocational education), only 30,12% achieved the highest grade.

Table 12: Distribution of pre-high school exit scores

Score	(1) classic	(2) scientific	(3) other
Ottimo	59,19	52,69	30,12
Distinto	26,82	30,75	28,37
Buono	12,70	14,23	30,68
Sufficiente	1,28	2,34	10,83
Total	100	100	100

Such a difference between classical and scientific tracks suggests the usefulness, for our analysis, of including the junior high school exit score in the vector of controls of our regressions. The index employed is called *std_score,* resulting from the standardization of grades with respect to the year of birth and the macro-area of origin (north-east, north-west, central, south and islands). This standardization is particularly necessary because in the central-southern regions, grades tend to be higher compared to the rest of the country (Brunello, Rocco, 2021).

Exploiting many other information contained in the PLUS survey, we construct in Table 13 a summary statistic where treatment and control groups are compared in terms of average characteristics. Individuals from classical studies are more likely to be females (69,1% vs 54,8%) and less likely to be grown in the north of the country (52,3% vs 58,4%); they have completed junior high school with larger grades (standardized score 0.319 vs 0.264) and tend to have a more privileged parental background (1.363 vs 0.92 ISEI-08 index).

Individuals with a classical background exhibit a larger probability of graduate from college (77,5% vs 72,7%) but a lower probability of being employed at the time of the questionnaire (68,1% vs 71,2%).

Employed individuals with a classical educational background earn, on average, higher annual real gross wages (log: 10.03 vs 9.98) and work on average a smaller number of annual hours (log: 7.32 vs 7.38); their average hourly real wage is slightly larger (2.7 vs 2.6).

Table 13: Summary Statistics – main variables

Variable	Classical	Scientific	Other
coll	.775	.727	.238
north	.523	.584	.607
female	.691	.548	.603
asil	.160	.159	.147
fs	.373	.366	.387
famback	1.363	.920	-.678
stdscore	.319	.264	-.117
emp	.681	.712	.620
training	.555	.536	.377
health	.790	.822	.753
lnhw	2.700	2.602	2.554
lnh	7.328	7.387	7.361
highocc	.506	.501	.287
lngw	10.028	9.989	9.916
partime	1.826	1.830	1.770
indet	.398	.419	.442
phappy	.596	.602	.620

3.4 The empirical approach

The main effort in attempting to address our research question involves seeking to solve the selection bias, which prevents the identification of the causal effect of interest (Angrist, Pischke, 2008).

Since the individual's cognitive ability affects both the future labour market outcomes and the high school choice, if abler individuals tend to choose the classical track over the scientific one, our estimates would be upward biased (Levine, Zimmerman, 1995). The literature has attempted to address this issue by looking for exogeneity conditions determined by policy rules, as in the Danish case mentioned in Chapter 2 (Joensen, 2006). Some authors (Silliman and

Virtanen, 2022) have estimated the labour market return to vocational secondary education using the Regression Discontinuity Design (RDD) method, leveraging admission criteria to high school and comparing individuals close to the admission threshold; this empirical approach, however, cannot be applied in the Italian case because enrolment in Italian high schools is free: students and their families choose the type of upper secondary education based on their interests and future prospects (Brunello, Rocco, 2023). Exceptions in this regard are represented by cases of excess demand, but admission criteria are not heterogeneous nationwide and may involve factors such as junior high school exit grades, distance and application timing (Brunello, Rocco, 2023)

Once the possibility of identifying an exogenous condition in the data is excluded, the empirical approach we employ is the Inverse Probability Weighted Regression Adjustment estimator (from this point forward, abbreviated as IPWRA).

IPWRA estimator relies on the Conditional Independence Assumption (CIA): conditional on the set of observable characteristics, the assignment to treatment is as good as random, or in other words, the individuals select themselves randomly into the classic or scientific lyceum (Angrist, Pischke, 2008).

The CIA, if satisfied, ensures that conditional on all observable characteristics, comparing individuals from the treatment and the control group means comparing "apples with apples": two individuals identical in terms of all the observable characteristics, differ only in terms of their high-school choice.

Let X be the vector of observable characteristics (which includes parental background, lower secondary education exit score, the region where an individual grew up before age 18 and so on); let define T {C, S} the set of treatments (classical lyceum diploma, scientific lyceum diploma): the Conditional Independence Assumption can be expressed as follows: $\{Y(C), Y(S)\} \perp C \mid X$. This assumption requires that all the variables affecting the educational choice are included into the vector of control X (there are no unobservable or omitted relevant characteristics) and that any residual variation in education is either random or due to factors that do not affect the outcomes of interest.

How does IPWRA estimator work? As already mentioned, for each individual we observe only the potential outcome associated to a particular treatment: we see Y(C) for individuals with classical studies, Y(S) for individuals with scientific studies.

Therefore, for individuals with a classical education, we fail to see the so-called counterfactual, which is Y(S), and vice versa for individuals with a scientific education.

IPWRA estimator imputes to each individual the missing potential outcomes, using information on similar peers (in terms of observable characteristics) who receive alternative treatments. It combines a treatment model that allows to predict the propensity score (probability of receiving the treatment) and another model to predict outcomes. IPWRA have the double-robust property: even if one of the models (treatment model or outcome model) is mis-specified, the estimator is still consistent (Angrist, Pischke, 2008).

More precisely, it consists in four steps (Angrist, Pischke, 2008):

1. Estimation of the probability of receiving the treatment variable, given the set of characteristics X.
2. Once the Propensity scores are estimated, inverse weights are computed. Each observation in the dataset is assigned an inverse weight based on their estimated probability of receiving the treatment. Individuals with low treatment probability receive high weights and vice versa.
3. The inverse probability weights computed are used to perform a weighted regression and obtain the potential outcomes associated to each value that the treatment variable can assume.
4. Computing the difference between average predicted outcomes, we obtain the Average Treatment Effects (ATEs). Our main ATE of interest is represented by the difference between the return to classical lyceum and the return to scientific lyceum: {E[Y(C)] – E[Y(S)]}.

3.5 The model

Building upon this strategy, we construct our research model. More specifically, we divide this section into three parts. In section 3.5.1, we elucidate the treatment variable that identifies each individual's high school choice. In section 3.5.2, we focus on the set of outcomes used to measure the high school return. Finally, in section 3.5.3, we describe the set of confounders used to address the Conditional Independence Assumption (CIA) upon which our identification strategy relies.

3.5.1 The treatment variable

In the regression model we use, the key variable is called *treatment*:

- Treatment =1 for all individuals who graduated from the classical lyceum
- Treatment = 2 for all individuals who graduated from the scientific lyceum

As depicted in Table 14, 34,59% of the sample holds a classical high school diploma, while 65,41% graduated from a scientific lyceum.

Table 14: Treatment variable

Treatment	Freq.	Percent	Cum.
1	5,291	34,59	34,59
2	10,007	65,41	100,00
TOTAL	15,298	100,00	

3.5.2 The outcomes of the model

In the context of this research, the returns associated with obtaining a high school education are assessed with regard to various dimensions, including educational achievements, labor market outcomes and quality of life indicators. To be more specific, the analysis involves the estimation of regression models that consider the following outcome variables:

- *Coll* is a binary variable, taking the value of 1 if the individual has attained at least a college degree, and 0 otherwise.
- *Emp* is a binary variable used to measure the likelihood of employment; it assumes a value of 1 if the individual has been employed in the past twelve months, relative to the time when the questionnaire was administered, and 0 otherwise.
- *Training* is a binary variable that takes the value of 1 if the individual has participated in seminars, conferences, training courses or professional development activities within the past 3 years, and 0 otherwise.
- *Health* is a binary variable used to gauge the self-perceived health status of the individual. It assumes a value of 1 if the individual rates his health as excellent or good and 0 if he considers it as satisfactory, mediocre, or poor.
- *Lnhw* represents the natural logarithm of hourly wages.
- *Lnh* represents the natural logarithm of annual hours worked.
- *Lngw* represents the natural logarithm of annual gross wage.

- *Highocc* is a binary variable used to assess the likelihood of working in a high-ranking occupation, such as managerial or professional roles.
- *Indet* is a binary variable designed to investigate whether the individual holds a permanent employment contract.
- *Partime* is another binary variable, distinguishing individuals with a part-time contract from those without it.
- *Phappy* is a binary variable employed to measure self-perceived happiness; it equals 1 if the individual perceives itself as happy and 0 otherwise.

Table 15 reported below illustrates the main summary statistic associated to each dependent variable.

Table 15: Outcomes – Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Emp	15,298	.701	.457	0	1
Training	15,298	.542	.498	0	1
Health	15,298	.811	.391	0	1
Phappy	15,298	.600	.489	0	1
Lnhw	7,920	2,634	.736	-3.049	7.20
Lnh	7,920	7.367	.391	3.871	8.841
Highocc	15,298	.503	.500	0	1
Lngw	7,920	10.002	.753	4.510	13.348
Partime	7,970	1.828	.388	1	2
Indet	15,298	.412	.492	0	1

For what concerns the variables Lnhw, Lnh, Lngw, we consider only data coming from employees, because earnings data related to self-employed individuals are not sufficiently reliable. We compute real hourly wages by dividing real gross annual earnings by the average number of annual hours worked.

3.5.3 The set of covariates

In line with the aforementioned discussion, the purpose of the control vector is to satisfy the Conditional Independence Assumption. Among the information extracted from the questionnaires PLUS (2014, 2016, 2018), all variables that influence both the individual's school choice and their future outcomes have been included in the set of controls. The list of confounders included in the model is the following:

- A fourth order polynomial in the year of birth (*yb, yb2, yb3, yb4*)
- A binary gender variable, with a value of 1 when the individual is female and 0 otherwise (*female*).
- A second-order polynomial in the parental background index introduced in Chapter 3.3 (*famback, famback2*)
- A second- order polynomial in the junior high school exit score, standardized by year of birth and macroarea (*stdscore, stdscore2*)
- A dummy variable with a value of 1 if the student, at age 14 (at the end of lower secondary education), had both parents living in his house, and 0 otherwise (*fs*)
- A dummy variable with a value of 1 if the respondent regularly attended a nursery school, and 0 otherwise (*asil*).

Furthermore, the two primary confounding variables, namely the parental background index and the cognitive ability index (*famback* and *stdscore*), are interacted with each other (*stdscore_fb*), with gender (*fam_fe, std_fe*), and with a dummy variable indicating whether the individual grew up in the northern regions of Italy (*fam_north, std_north*).

Taking into account the Italian scenario, it was deemed essential to incorporate geographical covariates and gender covariates into the control vector. As elucidated in Chapter 1, the classical lyceum remains attractive in the southern regions, with a higher proportion of students choosing it. Simultaneously, we expect the geographical component to be correlated with most of the dependent variables under consideration.

Current and historical macroeconomic data indicate that Northern Italy, characterized by a well-developed industrial base, exhibits higher per capita income and lower unemployment rates (De Philippis, Locatelli, 2022). In other word, growing up in Southern Italy may be negatively correlated with labour market income, albeit mitigated by the phenomenon of internal migration towards the Northern regions, a trend that has characterized our country for decades.

Similarly, the same line of reasoning prompted us to control for gender. As specified in Chapter 1, females tend to choose the classical high school more frequently than their male counterparts. Additionally, in Italy, the female labor force participation rate is lower than that of males, and the labour market tends to offer, on average, lower wages to women compared to men, even under similar conditions (Zizza, 2013).

3.6 What is the probability of choosing the classical lyceum?

After drawing the model, we are now using our estimator to investigate the high school effect of interest.

As seen in section 3.4, the first step in applying the IPWRA estimator involves estimating the probability of receiving the treatment, given the set of characteristics X. To do this, we use a logistic model (LOGIT) and estimate the probability of enrolling in the classical lyceum as a function of the covariates available in our dataset.

Table 16 illustrates the marginal effects that each covariate has on the probability of receiving the treatment: the results must be interpreted with respect to the probability of receiving the control, that is, enrolling in the scientific lyceum.

Being born in Northern Italy reduces the probability of choosing classical studies by 5,18%, and the effect is statistically significant. As expected, being female increases the probability by 14,55%, while having attended a nursery school has a positive and statistically significant effect (+4,4%).

Moreover, coming from a more privileged family socio-economic background increases the probability of choosing classical studies over scientific ones (.0165/.005) while, contrary to our hypothesis, the pre-enrollment student's cognitive ability does not seem to influence the choice in a statistically significant manner.

Table 16: Probability of choosing the classical lyceum: logistic model

	Marginal effects
yb	-.027 (.023)
yb2	.141 (.169)
yb3	-.034 (.050)
yb4	.002 (.005)
north	-.0518 (.010)***
Female	.1455 (.010)***
Asil	.0440 (.013)***
FS	.004 (.019)
Famback	.0165 (.005)***
Famback 2	.0034 (.000)***
Stdscore	.0168 (.014)
Stdscore2	-.003 (.008)
Stdscore-fb	.000 (.002)
Fam_north	-.004 (.003)
Std_north	-.002 (.014)
Fam_Fe	.000 (.003)
Std_Fe	.010 (.014)
Scoremiss	.028 (.025)
Famback_Miss	.008 (.035)
Finescuola_Miss	-.010 (.030)

Statistically significant at the 10% level / **Statistically significant at the 5% level / * Statistically significant at the 1% level.*

Note – Standard errors are reported in parentheses. In order to enhance the interpretability of the results, the current table includes the "North" variable in place of the regional covariates used in STATA (rb2-rb18); this choice does not alter the magnitude or significance of the results.

3.7 The results

We structure the exposition of our findings as follows. Section 3.7.1 offers the baseline estimates. Section 3.7.2, on the other hand, is dedicated to the comparison of high school returns among different geographical areas within Italy. In Section 3.7.3, we assess the high school return by gender. Lastly, Section 3.7.4 examines the college return by high school type.

3.7.1 Return to high school

The primary goal of this paper is to seek an understanding of whether the type of high school chosen has an impact on future educational attainments and labour market outcomes. More specifically, given that existing literature has primarily focused on comparing vocational and academic education, we have chosen to concentrate on different types of academic education, by comparing classical and scientific studies.

Table 17 illustrates the estimates of the average potential outcomes by treatment, which are used to compute the difference between returns to high school, presented in Table 18.

In our sample, the probability of graduating for a student coming from a classical lyceum is 67,4%, percentage that falls to 63,9% for a student coming from a scientific lyceum. The difference, amounting to 3,52%, is positive and statistically significant. This result aligns with the findings in the paper "Do Classical Studies open your mind?" (Brunello, Rocco, 2023): those who have completed classical studies are more likely to graduate from college.

As shown in Table 18, having a classical diploma reduces the employment probability by 2,48% compared to a scientific diploma; the difference is statistically significant at a 1% level.

Furthermore, we find that classical studies have a positive, although weakly significant (90% confidence interval), effect on the training probability: they increase by 1,98% the likelihood of having participated in seminars, conferences and training courses in the last three years.

When we consider self-perceived health status and self-perceived happiness, we find that the return to high school does not significantly differ between the two groups.

Table 17: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

	Classic	Scientific
	A. full sample	
College	.674***	.639***
	(.009)	(.006)
Employment probability	.788***	.813***
	(.006)	(.004)
Training probability	.573***	.553***
	(.009)	(.006)
Self-perceived Health status	.795***	.807***
	(.007)	(.004)
Self-perceived happiness	.867***	.860***
	(.007)	(.005)
	B Employees only	
Log hourly wages	2.668***	2.648***
	(.016)	(.010)
Log annual hours worked	7.359***	7.398***
	(.009)	(.005)
High-ranked occupation	.680***	.682***
	(.010)	(.006)
Log annual gross wage	10.02***	10.04***
	(.016)	(.010)
Part time	.122***	.116***
	(.006)	(.004)
Indet	.566***	.606***
	(.010)	(.007)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

Looking at labour market outcomes and focusing solely on the subsample of employees, we find that having attended a classical lyceum reduces the number of annual hours worked by 3,94%, while no statistically significant differences have been found in terms of hourly wage and annual gross wage (see Table 18). One possible interpretation could be linked to the fact that classical studies, being more associated with public sector jobs, lead to a lighter working schedule.

When considering the probability of having a permanent contract, we find that the return to high school is lower for classical students compared to scientific ones. The difference shown in Table 18, amounting to 3,99%, is negative and statistically significant.

Finally, among employed individuals, the return to high school in terms of high-ranked occupation does not exhibit significant differences between treatment and control groups.

Overall, the presented results show that attending a classical high school yields a higher return in terms of educational attainments. However, this return does not seem to be reflected in better labor market outcomes. On the contrary, it appears that classical studies, in comparing to scientific studies, generate a “precarity effect”, reducing the likelihood of being employed and obtaining a permanent contract.

Table 18: Differences between Returns to high-school

E [Y (C) – Y (S)] Classic vs Scientific Lyceum

(Estimation Method: IPWRA)

	Difference between Returns to High School
	A. FULL SAMPLE
College	.0352***
	(.010)
Employment probability	-.0248***
	(.007)
Training probability	.0198*
	(.011)
Self-perceived health status	-.012
	(.008)
Self-perceived happyness	.006
	(.009)
	B. Employees only
Log hourly wages	.020
	(.018)
Log annual hours worked	-.0394***
	(.011)
High-ranked occupation	-.002
	(.012)
Log annual gross wage	-.018
	(.018)
Part time	.005
	(.007)
Indet	-.0399***
	(.012)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

3.7.2. Return to high school by Area

For a more detailed analysis, we also present estimates of the return to high school differences by geographical areas. In particular, Table 19 estimates the average potential outcomes by treatment in the sub-sample consisting of individuals who grew up in Northern Italy until the age of 18, while Table 20 estimates the average potential outcomes by treatment in the residual sub-sample composed of individuals who grew up in Central and Southern Italy (including the Islands, Sicily and Sardinia).

The results obtained, as illustrated in Table 21, are not easily interpretable. First and foremost, we cannot disregard the fact that Italy, from the post-war period to the present day, has experienced significant and impactful internal migration from the South to the North of the country (Bonifazi, 2020). This implies that a portion of our sample grew up in Southern regions but subsequently migrated to the North, where they predominantly pursued their career.

Secondly, with regard to the variables for which we only consider employed works, the North/Central-South division results in a small sample size for each subgroup, making the obtained results less reliable. In more detail, the Northern Italy sample comprises 8.616 individuals, but for some variables (log hourly wages, log annual hours worked, log annual gross wage) it reduces to 4.963 units; the Central-South Italy sample consists of 6.682 observations, but for the same aforementioned outcomes, it reduces to 2.957 units.

By comparing the average potential outcomes in the two geographical areas (see Table 19 and Table 20), we can observe the presence of territorial differences in terms of the underlying economic landscape. Specifically, the employment probability, which indicates whether the interviewee has been employed in the last 12 months relative to the questionnaire administration, is above 80% in Northern Italy (for both the treatment and control groups) and slightly higher than 70% in the Central-South.

In Table 21, we present estimates of the differences in returns to high school in the two reference geographical areas. Among those who were born and raised in Northern Italy, graduating from the classical lyceum reduces the probability of employment by 3,85%, while this negative and statistically significant effect disappears in the Central-South sub-sample.

Let’s attempt to interpret these results based on territorial economic differences: the Central-Southern regions are characterized by a large public sector (Brunello, Rocco, 2021) with a higher relative number of public employees, working in regional and provincial administrations, healthcare, and education. In contrast, Northern Italy is characterized by a significant industrial sector.

Given these aspects, it is realistic to hypothesize that scientific studies are complementary to the industrial structure of Northern Italy, as the latter raises the demand for workers with technical and scientific education. Conversely, classical studies may be more complementary to the economic network of Central-Southern Italy, thus mitigating the negative employment effect seen above.

The other two significant differences we find in Table 21 pertain to the variables *training* and *Indet.* In the Central-South, attending the classical lyceum increases the likelihood of participating in training and development courses by 4,23%, while in Northern Italy the difference becomes close to zero and not statistically significant.

When considering the probability of permanent employment, the precariousness effect of classical studies vanishes in the Central-South; conversely, in Northern Italy, classical studies reduce this probability by 4,7% compared to scientific studies.

Table 19: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

North of Italy

	Classic	Scientific
	A. full sample	
College	.680***	.645***
	(.012)	(.008)
Employment probability	.837***	.875***
	(.007)	(.004)
Training probability	.578***	.572***
	(.012)	(.008)
Self-perceived Health status	.812***	.826***
	(.009)	(.006)
Self-perceived happiness	.880***	.868***
	(.009)	(.007)
	B Employees only	
Log hourly wages	2.65***	2.64***
	(.019)	(.011)
Log annual hours worked	7.37***	7.41***
	(.012)	(.006)
High-ranked occupation	.663***	.662***
	(.013)	(.008)
Log annual gross wage	10.03***	10.06***
	(.019)	(.011)
Part time	.129***	.122***
	(.008)	(.005)
Indet	.585***	.632***
	(.014)	(.008)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

Table 20: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

CENTRE-SOUTH of Italy

	Classic	Scientific
	A. full sample	
College	.667***	.629***
	(.012)	(.009)
Employment probability	.716***	.721***
	(.010)	(.007)
Training probability	.567***	.524***
	(.012)	(.009)
Self-perceived Health status	.772***	.778***
	(.010)	(.008)
Self-perceived happiness	.848***	.849***
	(.011)	(.008)
	B Employees only	
Log hourly wages	2.69***	2.65***
	(.028)	(.018)
Log annual hours worked	7.32***	7.36***
	(.014)	(.010)
High-ranked occupation	.713***	.718***
	(.015)	(.010)
Log annual gross wage	10.02***	10.01***
	(.028)	(.019)
Part time	.109***	.106***
	(.009)	(.006)
Indet	.531***	.559***
	(.015)	(.011)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

Table 21: Differences between Returns to High School, by Area (Estimation Method: IPWRA)

	Difference between returns to High School	
	A. NORTH	B. CENTER AND SOUTH
College	.0346**	.0382**
	(.014)	(.015)
Employment probability	-.0385***	-.005
	(.008)	(.012)
Training probability	.005	.0423***
	(.014)	(.015)
Self-perceived Health status	-.014	-.005
	(.011)	(.013)
Self-perceived happiness	.011	-.000
	(.011)	(.014)
Log hourly wages	.004	.048
	(.022)	(.033)
Log annual hours worked	-.0369***	-.0435***
	(.014)	(.017)
High-ranked occupation	.001	-.005
	(.016)	(.018)
Log annual gross wage	-.032	.004
	(.022)	(.033)
Part time	.007	.003
	(.010)	(.011)
Indet	-.047***	-.027
	(.016)	(.019)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

3.7.3 Return to high school by gender

Table 24 presents the gender – based disparities in the returns to high school education. The female sub-sample comprises 9.141 individuals, while the male sub-sample consists of 6.157 observations. The estimates are derived as the difference from the Average Potential Outcomes associated with each treatment status. For further details, Tables 22 and 23 display the Average Potential Outcomes by treatment and by gender, obtained using the IPWRA estimator.

Upon comparing the results, we find that having attended a classical high school increases the likelihood of obtaining a college degree for women by 3,86%, while the effect becomes not statistically significant in the male population. Regarding the return to high school in terms of employment probability, the negative and statistically significant effect of classical studies is more pronounced in absolute value within the male sample (-3,52%) compared to the female sample (-1,99%).

Concerning the return to high school in terms of training probability, classical studies are associated with a higher probability for women (+2,55%), while this positive effect diminishes in the male population.

Examining the *Indet* variable, we observe that the “job insecurity” effect induced by a classical education is absent in the male sample, whereas among women holding a classical high school diploma reduces the probability of having a permanent contract by 5,33%.

The estimates related to the likelihood of working in a high-ranking occupation (managers, professionals, white collar) show a sign reversal between the male and female populations. In particular, classical studies seem to increase this probability for women (+3,54%) and decrease it for men (-3,63%). As in the case of the geographic analysis, the small sample size in this context makes these coefficients challenging to interpret causally.

Lastly, we find that having pursued classical studies reduces the probability of women perceiving themselves as happy (-2,04%), while it increases this probability for men (+3,67%). Both effects are statistically significant within a 95% confidence interval. One possible interpretation of this result could be linked to the fact that in the past, female graduate of classical lyceum, after obtaining their high school diploma and degree, often pursued teaching careers. Such a career path might generate personal dissatisfaction related to low-income levels and limited opportunities for professional growth. Nevertheless, these considerations remain hypotheses that require further empirical support.

Table 22: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

Female

	Classic	Scientific
	A. full sample	
College	.714***	.676***
	(.009)	(.007)
Employment probability	.740***	.760***
	(.007)	(.006)
Training probability	.560***	.534***
	(.009)	(.007)
Self-perceived Health status	.783***	.800***
	(.007)	(.006)
Self-perceived happiness	.864***	.884***
	(.008)	(.006)
	B Employees only	
Log hourly wages	2.65***	2.63***
	(.016)	(.012)
Log annual hours worked	7.27***	7.31***
	(.011)	(.008)
High-ranked occupation	.700***	.664***
	(.010)	(.009)
Log annual gross wage	9.935***	9.948***
	(.017)	(.012)
Part time	.170***	.178***
	(.008)	(.006)
Indet	.597***	.650***
	(.011)	(.008)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

Table 23: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

Male

	Classic	Scientific
	A. full sample	
College	.630***	.599***
	(.016)	(.009)
Employment probability	.840***	.875***
	(.009)	(.004)
Training probability	.586***	.575***
	(.016)	(.009)
Self-perceived Health status	.809***	.816***
	(.012)	(.007)
Self-perceived happiness	.871***	.834***
	(.012)	(.009)
	B Employees only	
Log hourly wages	2.69***	2.66***
	(.030)	(.016)
Log annual hours worked	7.44***	7.49***
	(.017)	(.007)
High-ranked occupation	.664***	.700***
	(.017)	(.009)
Log annual gross wage	10.13***	10.16***
	(.028)	(.016)
Part time	.073***	.056***
	(.009)	(.004)
Indet	.537***	.563***
	(.018)	(.010)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

Table 24: Differences between Returns to high school by gender

	A. Female	A. Male
College	.0386***	.0306
	(.011)	(.019)
Employment probability	-.0199**	-.0352***
	(.009)	(.010)
Training probability	.0255**	.010
	(.012)	(.018)
Self-perceived Health status	-.0162*	-.006
	(.009)	(.014)
Self-perceived happiness	-.0204**	.0367**
	(.010)	(.015)
Log hourly wages	.0214	.0271
	(.019)	(.033)
Log annual hours worked	-.0346***	-.051***
	(.013)	(.019)
High-ranked occupation	.0354***	-.0363*
	(.014)	(.020)
Log annual gross wage	-.013	-.0243
	(.020)	(.031)
Part time	-.008	.016
	(.011)	(.010)
Indet	-.0533***	-.025
	(.014)	(.021)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

3.7.4 Return to college by high school type

The results presented so far demonstrate that, despite a higher likelihood of obtaining a college degree, classical studies are associated with worse labour market outcomes compared to their counterparts who have completed a scientific high school. It seems that, particularly in terms of employment probability and the likelihood of securing a permanent contract, scientific studies generate a greater return compared to classical studies.

Through what processes is this advantage determined?

One initial hypothesis is that scientific studies provide students with skills and knowledge that better align with the demands of the labour market, and this effect occurs in a direct manner.

A second hypothesis is that the scientific lyceum is more complementary to college education compared to the classical lyceum. Beyond the fact that classical studies increase the probability of an individual attaining a college degree, we want to examine whether the return to college, in terms of economic outcomes, is greater for scientific students compared to their classical counterparts.

In order to estimate the return to college by high school type, we assume that individuals can choose different combinations of high school type and college: in order to do this, we generate a variable called *Treat*, such that:

- Treat = 1 for college graduates with classical studies (CC: 26,81% of the sample)
- Treat = 2 for individuals with a classical high school diploma, without a college degree (CNC: 7,77% of the sample)
- Treat = 3 for college graduates with scientific studies (SC: 47,59% of the sample)
- Treat = 4 for individuals with a scientific high school diploma, without a college degree (SNC: 17,83% of the sample)

Let T = {CC, CNC, SC, SNC} be the set of treatment described above; let {Y(CC), Y(CNC), Y(SC), Y(SNC)} be the set of potential outcomes, each of one is associated to one of the four treatment groups. Given the set of covariates X already used to compute the returns to high school, we apply the IPWRA estimator to find:

- The return to college for individuals with a classical high school diploma E [Y(CC) – Y(CNC)]
- The return to college for individuals with a scientific high school diploma E [Y(SC) – Y (SNC]
- Differences between returns to college by high school type, computed as E [Y(CC) – Y(CNC)] - E [Y(SC) – Y (SNC].

Results are presented in Table 25 and Table 26. In particular, Table 25 shows the average potential outcomes associated to the four values that the variable *Treat* can assume, while Table 26 illustrates if the returns to college differ significantly between individuals with classical and scientific studies.

Table 25 -Return to college by high school type

Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

	A. CC	B. CNC	C. SC	D. SNC
		A. full sample		
Employment probability	.830***	.679***	.859***	.716***
	(.005)	(.014)	(.003)	(.009)
Training probability	.657***	.387***	.636***	.395***
	(.009)	(.018)	(.006)	(.012)
Self-perceived	.809***	.767***	.832***	.757***
Health status	(.007)	(.014)	(.005)	.010)
Self-perceived happiness	.867***	.865***	.874***	.838***
	(.009)	(.012)	(.006)	(.010)
		B Employees only		
Log hourly wages	2.72***	2.47***	2.72***	2.48***
	(.020)	(.027)	(.012)	(.017)
Log annual hours worked	7.36***	7.34***	7.39***	7.40***
	(.011)	(.018)	(.006)	(.012)
High-ranked occupation	.791***	.398***	.797***	.425***
	(.010)	(.022)	(.006)	(.015)
Log annual gross wage	10.09***	9.82***	10.11 ***	9.88***
	(.020)	(.029)	(.012)	(.020)
Part time	.096***	.197***	.093***	.180***
	(.006)	(.016)	(.004)	(.011)
Indet	.561***	.555***	.593***	.621***
	(.011)	(.023)	(.007)	(.014)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

We find that, for all the economic and labour market dependent variables taken into account, the difference between returns to college among the two groups is very small and not statistically significant (see Table 26). For example, graduate from a college increases the employment probability by 15,1% among individuals with classical studies (83% vs 67,9%, see Table 25), and by 14,3% among individuals with a scientific high school diploma (85,9% vs 71,6%, see Table 25).

Interestingly, scientific studies seem to make students more appealing to the labour market even in the absence of a degree. The employment probability for an individual belonging to the group called SNC (Scientific high school diploma, No College Degree) is 71.6%, with respect to 67.9% for the group called CNC (Classical high school diploma, No College Degree).

For what concerns the self-perceived health status, we find that having a college degree increases the probability of perceive a positive health status, but this return is smaller for individuals with a classical education: the difference (3,36%) is statistically significant at a 90% level of confidence.

Similarly, also the return to college in terms of self-perceived happiness differ between the two groups: for classical students the return is smaller (-3,35%) and the difference is statistically significant at a confidence interval of 90%.

Overall, we conclude that there is no evidence of strong complementarities between scientific education and college degree, with respect to classical education and college degree.

TABLE 26: Differences between Returns to college

E[Y (CC) – Y (CNC)] - E[Y (SC) – Y (SNC)] Classic vs Scientific Lyceum

(Estimation Method: IPWRA)

	Difference between Returns to college
	A. FULL SAMPLE
Employment probability	.007
	(.018)
Training probability	.028
	(.025)
Self-perceived health status	-.0336*
	(.020)
Self-perceived happyness	-.0335*
	(.020)
	B. Employees only
Log hourly wages	.010
	(.039)
Log annual hours worked	.025
	(.025)
High-ranked occupation	.020
	(.030)
Log annual gross wage	.035
	(.042)
Part time	-.014
	(.021)
Indet	.035
	(.031)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

3.8 Overlap condition

The causal interpretation of our estimates relies on two key assumptions:

- The Conditional Independence Assumption (CIA) posits that, given our set of control variables, the assignment of an individual to either the classical lyceum (Treatment=1) or the scientific lyceum (Treatment = 2) is independent of the outcomes. This means that, once we have considered the control variables, there should be no systematic association between the high school choice and the outcome variables. The CIA is crucial for ensuring that the estimated treatment effects are unbiased and accurately reflect the causal relationship we are investigating.
- The Overlap Assumption states that each individual has a non-negative probability of receiving each treatment level. It is essential for the construction of valid inverse probability weights, and it ensures the estimation to proceed without encountering issues related to division by zero (Rubin, 2005)

For what concerns the first assumption, CIA cannot be tested. However, we have confidence that the rich set of covariates employed can minimize the risk of omitting relevant factors that affect the high school choice and the outcomes of our model.

Let's now focus on the second assumption. In order to interpret our estimates as causal effects, we need to verify that the overlap condition holds.

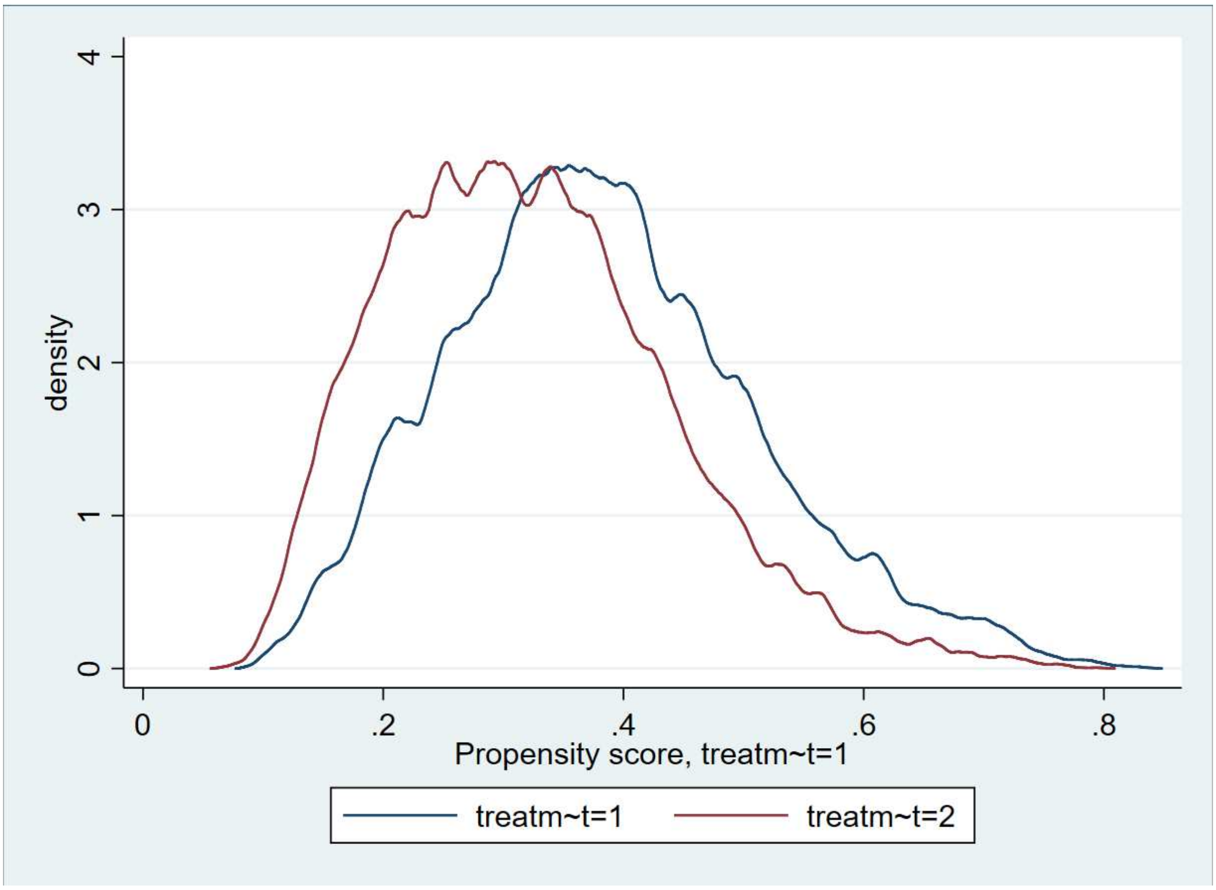
Figure 2 plots the estimated densities of the probability of choose the classical lyceum, conditional on the set of characteristics X.

The distribution of propensity scores in the two curricula substantially overlaps.

In other words, conditioned on the set of characteristics X, the probability of receiving the treatment (i.e. classical lyceum) is substantially the same, ex-ante, between the group of those who will actually choose the classical lyceum and their peers who will instead choose the scientific lyceum.

This result is not surprising: even before controlling for the effect of confounders, the two groups appear similar to each other as they are drawn from the upper part of the population distribution in terms of family background and cognitive ability. The proper use of control variables, as evident in Figure 2, has made the two distributions overlapped.

Figure 2: **Propensity score by High school type – Overlap condition.**



Furthermore, in order to assess the ability of the IPWRA estimator to rebalance the sample through the use of inverse probability weights, we compare covariate balancing between the treatment group and the control group. Table 27 presents the standardized differences and variance ratios for each control variable included in our model. In particular, it compares the raw data and the weighted data in different columns to assess the ability of the weights to rebalance the sample: balancing is considered satisfactory if, after weighting, the standardized differences approximate 0, and the variance ratios are close to 1.

As shown in Table 27, for most of the variables, covariance balancing can be considered satisfactory. For example, let's consider two variables that, as shown in Chapter 3.6.1, have a statistically significant effect on the probability of choosing the classical lyceum: *female* and *famback*. For the gender index, the standardized differences are -0.298 in the raw data and 0.008 (close to 0) in the weighted data, while the variance ratios are 1.160 in the raw data and 0.996 (close to 1) in the weighted data. For the parental background index, the standardized differences are -0.179 in the raw data and 0.025 (close to 0) in the weighted data, while the variance ratios are 0.837 in the raw data and 1.010 (close to 1) in the weighted data.

Table 27: Covariate balancing: Scientific lyceum relative to Classical Lyceum

Variable	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
yb	.174	-.004	.996	.993
yb2	.184	-.006	1.062	.997
yb3	.190	-.006	1.113	.999
yb4	.192	-.006	1.153	.999
rb2	.110	.010	1.303	1.024
rb3	.038	.010	1.328	1.081
rb4	.020	-.015	1.084	.939
rb5	.022	-.005	1.155	.968
rb6	-.008	-.011	.956	.943
rb7	.054	-.007	1.270	.968
rb8	.038	-.016	1.162	.940
rb9	-.026	-.015	.850	.908
rb10	.029	-.015	1.186	.916
rb11	-.085	.006	.798	1.019
rb12	.026	.011	1.127	1.053
rb13	-.037	-.009	.915	.977
rb14	-.024	.009	.918	1.033
rb15	-.016	-.002	.892	.983
rb16	-.018	.013	.931	1.054
rb17	-.106	.003	.748	1.009
rb18	-.048	.000	.790	1.002
female	-.298	.008	1.160	.996
asil	-.003	.035	.993	1.068
fs	-.015	-.003	.991	.998
famback	-.179	.025	.837	1.010
famback2	-.217	.020	.714	1.084
stdscore	-.080	.011	1.096	.995
stdscore2	.015	.003	1.232	1.078
stdscore_fb	-.112	.000	.780	1.025
fam_north	-.064	.018	.858	.964
std_north	-.010	.007	1.178	1.006
fam_fe	-.223	.021	.643	1.044
std_fe	-.127	.019	.835	.967
scoremiss	-.029	-.011	.980	.992
famback_miss	.022	-.010	1.105	.957
finescuola_miss	-.013	-.000	.988	.999

Note – Standardized differences are computed as the ratio of average differences to the squared root of the sum of variances.

3.9 Sensitivity analysis

As specified thus far, our research question focuses on the comparison between two curricula: classical and scientific. For this reason, we constructed a sample of 15.298 individuals, of which 5.291 hold a classical diploma and 10.007 hold a scientific diploma. However, the PLUS dataset from which we derived the data allows us to enrich the sample by including individuals who have obtained a high school diploma different from both classical and scientific curricula.

We have already taken advantage of this opportunity in Chapter 3.3, which pertains to the statistical description of the data. In that context, the comparison with the remaining part of the student population allowed us to assert that classical and scientific students are drawn from the upper part of the population distribution in terms of maternal and paternal education, paternal occupation, and the pre-enrolment pupil's cognitive ability.

Increasing the sample size in econometric analyses offers several advantages (Stock, Watson, 2007), including:

- A larger sample provides greater statistical power, enabling a more detailed and specific analysis, capable of detecting in some cases subtler differences between the treatment and control groups.
- The larger the sample, the more representative it is of the general population, making the research results more generalizable.
- As the sample size increases, statistical estimates tend to converge towards the true population values, reducing uncertainty and statistical error.

For these reasons, we decided to conduct a sensitivity analysis, investigating whether the results obtained in terms of differences in high school returns are confirmed even when using a sample enriched with observations of individuals with a high school diploma different from classical or scientific.

The Table 28 below presents Question D88_1 from the PLUS questionnaire (2014,2016,2018), which provides the necessary information for classifying students based on their high school diploma.

Table 28: high school diploma

D88_1. Quale diploma o licenza liceale ha conseguito?	Freq.	Percent	Cum.
Diploma di qualifica professionale	2.790	5,40	5,40
IPS, Istituto d'Arte, Scuola Magistrale	887	1,72	7,12
Liceo classico	5.291	10,24	17,36
Liceo scientifico	10.007	19,38	36,74
Liceo Linguistico, artistico e altri tipi	2.673	5,18	41,91
Istituto Professionale (di 5 anni)	3.988	7,72	49,64
Istituto tecnico (Agrario, Industriale, Nautico, Aeronautico etc.)	10.437	20,21	69,84
Ragioneria e Geometri	10.513	20,35	90,20
Istituto Magistrale – Liceo Psico/Socio pedagogico	4.841	9,37	99,57
Conservatori / Accademie	222	0,43	100,00
TOTAL	51.649	100,00	

Using the information contained in Table 28, we construct the new *treatment* variable, which can take on three different values:

- Treatment = 1 if the individual has a classical high school diploma
- Treatment = 2 if the individual has a scientific high school diploma
- Treatment = 3 if the individual has a different type of high school diploma, neither classical nor scientific.

Table 29 displays the breakdown of our sample based on the treatment variable. The sample size increases to 51.649 individuals, with 10,24% holding a classical diploma, 19,38% having a scientific diploma, and the third group, labelled " Other", representing the remaining portion of the sample, accounting for 70,38%.

Table 29: Treatment variable

Treatment	Freq.	Percent	Cum.
1	5,291	10,24	10,24
2	10,007	19,38	29,62
3	36,351	70,38	100,00
TOTAL	51,649	100,00	

The majority of individuals belonging to the third treatment group has a vocational education (predominantly from "Istituti tecnici e professionali"). However, it is important to specify that a portion of this group consists of individuals with academic diplomas, such as linguistic, artistic, and socio-pedagogical diplomas. This grouping is due to the fact that the primary focus of our research is the comparison between classical and scientific licei.

Symmetrically to what observed in Chapter 3.6.2, we present the results divided into two tables: Table 30 illustrates the estimates of the average potential outcomes by treatment, while Table 31 shows the differences between returns to high school.

Firstly, when observing the average potential outcomes (see Table 30), it is easy to note that the main differences do not concern classical and scientific licei but rather the comparison between academic and vocational education.

For example, the probability of graduating is 65,8% for an individual with a classical diploma, 60,9% for one with a scientific education, and only 21,8% for an individual belonging to the "*Other*" group. Smaller average potential outcomes are also observed in terms of employment probability, hourly wages, and annual gross wages.

Regarding the probability of having a high-ranked occupation, in the *Other* group, it averages only 48,2% compared to approximately 67% in classical and scientific diploma holders.

Resuming the comparison between classical and scientific, Table 31 confirms that the differences in returns to high school are not statistically significant in terms of self-perceived health status, self-perceived happiness, hourly wages, annual gross wages, probability of having a high-ranked occupational status, and probability of having a part-time contract.

Having a classical high school diploma increases the likelihood of achieving a college degree by 4,9% compared to having a scientific diploma. However, it diminishes the probability of employment and the likelihood of securing a permanent job contract by 2,29% and 4,61% respectively.

The positive impact of a classical education on vocational training appears to exhibit weak statistical significance. Additionally, classical studies result in a 5,55% reduction in the annual working hours compared to scientific studies.

In summary, the sensitivity analysis corroborates our findings. Despite the fact that classical studies positively influence educational achievements, they exert adverse effects on labour market outcomes, notably in terms of employability and the likelihood of securing a stable employment contract.

Table 30: Average Potential Outcomes by Treatment (Estimation Method: IPWRA)

	Classic	Scientific	Other
	A. full sample		
College	.658***	.609***	.218***
	(.010)	(.007)	(.002)
Employment probability	.785***	.808***	.754***
	(.007)	(.004)	(.002)
Training probability	.567***	.537***	.430***
	(.010)	(.007)	(.003)
Self-perceived	.777***	.790***	.758***
Health status	(.009)	(.006)	.003)
Self-perceived happiness	.867***	.863***	.882***
	(.008)	(.006)	(.003)
	B Employees only		
Log hourly wages	2.69***	2.66***	2.57***
	(.019)	(.010)	(.006)
Log annual hours worked	7.33***	7.39***	7.37***
	(.012)	(.006)	(.003)
High-ranked occupation	.678***	.674***	.482***
	(.012)	(.007)	(.004)
Log annual gross wage	10.033***	10.054***	9.94***
	(.020)	(.011)	(.006)
Part time	.133***	.125***	.169***
	(.008)	(.005)	(.003)
Indet	.587***	.633***	.689***
	(.012)	(.007)	(.004)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

TABLE 31

Differences between Returns to high-school

E[y (c) – y (s)]   Classic vs Scientific Lyceum

(Estimation Method: IPWRA)

	Difference between Returns to High School
	A. FULL SAMPLE
College	.0490***
	(.013)
Employment probability	-.0229***
	(.008)
Training probability	.0301**
	(.012)
Self-perceived health status	-.0131
	(.010)
Self-perceived happyness	.003
	(.010)
	B. Employees only
Log hourly wages	.0346
	(.022)
Log annual hours worked	-.0555***
	(.013)
High-ranked occupation	.003
	(.014)
Log annual gross wage	-.020
	(.022)
Part time	.008
	(.010)
Indet	-.0461***
	(.015)

**Statistically significant at the 10% level / **Statistically significant at a 5% level / ***Statistically significant at a 1% level*

Note – Standard errors are in parentheses.

CONCLUSIONS

In this research work, we attempted to make an empirical contribution to the ongoing debate in Italy regarding the usefulness of the classical lyceum.

The classical lyceum is a historical Italian institution that, after representing the elite high school for decades, is currently experiencing a significant decline in enrollments, particularly in Northern Italy.

While proponents of the classical lyceum emphasize its ability to foster critical thinking in students, critics, on the hand, draw attention to the time and energy spent for the so-called "dead languages" (latin and ancient Greek), especially in light of a labor market increasingly oriented toward mathematical and scientific skills.

Our paper also contributes to the international literature studying the economic impact of high school curricula, with a specific focus on differences within academic education, through the comparison of classical and scientific studies.

Our empirical analysis exploits data drown from PLUS (Participation, Labour, Unemployment Suvey), an Italian survey which contains information on students 'educational pathways.

The ideal empirical approach would have been to simulate a natural experiment by identifying an exogenous variation in the data induced by a policy rule. However, this strategy is hindered by the fact that in Italy enrolment in upper secondary education is free, and admission criteria in cases of excess demand vary by geographical areas. For this reason, we decide to adopt a methodology based on the Inverse Probability Weighted Regression Adjustment estimator (IPWRA), using a rich set of control variables capable of capturing the primary factors affecting the high school choice, such as the family background and the individual's pre-enrollment cognitive ability.

We compare the return to classical lyceum and the return to scientific lyceum by considering the following set of outcomes: probability of graduating from college, employment probability, training probability, self-perceived health status, self-perceived happiness, real hourly wage, annual hours worked, real annual gross wage, probability of working in a high-ranking occupation, probability of holding a permanent employment contract and probability of having a part-time contract.

We find that attending a classical lyceum yields a higher return in terms of educational attainments: an individual with a classical background is 3,52% more likely to graduate from college with respect to a peer with a scientific high school diploma. However, this return does not seem to be reflected in better labor market outcomes. We show that pursuing classical education, when compared to a scientific curriculum, results in a precariousness effect. In particular, in the baseline estimates we find that having a classical high school diploma reduces the likelihood of being employed by 2,48%, and the probability of securing a permanent job contract by 3,99%. Moreover, we also find that classical studies reduce the numbers of annual hours worked by 3,94%, without any significant effect on real hourly wage and real annual gross wage.

When we split the sample between Northern and Central/Southern regions, we discover that the labour market penalty induced by classical studies remains statistically significant only in Northern Italy but disappears in the remaining part of the Country. We attempted to causally interpret these results based on the economic differences across different geographic areas: scientific studies align with the industrial network of Northern Italy, given its increased need for technically and scientifically educated workers. On the other hand, classical studies may exhibit greater synergy with the economic framework of Central-Southern Italy, characterized by a large public sector.

Finally, we questioned whether the larger labour market outcomes associated with a scientific high school education could be attributed to a greater complementarity with university studies, resulting in a higher return to college. However, we find that the differences in returns to college between individuals with a classical background and those with a scientific background are very small and not statistically significant. Thus, a second hypothesis remains, suggesting that scientific studies provide individuals skills more marketable beyond their post-secondary education path.

The results of this analysis should be interpreted with caution, as it presents some limitations.

First and foremost, our dataset does not allow us to estimate the high school return in terms of lifetime labour market outcomes; instead, it captures these variables at a specific point in an individual's working life.

Second, our comparison between classical and scientific track considers individuals from different cohorts who graduated in different years, during which the curricula of the two Italian institutions changed.

Finally, we cannot guarantee that we have completely ruled out the presence of selection bias that may affect the causal interpretation of our results. In particular, the lower-secondary education exit grade serves as only an approximation of the individual's pre-enrollment cognitive ability, which, as previously indicated in the literature, remains unobservable.

However, aware that our considerations and results require further empirical support, we hope to stimulate the debate on the classical lyceum and how its decline can be addressed.

APPENDIX

In the following, a description of the steps leading to the construction of the *famback* variable is provided, where *famback* represents our parental background index. This index is constructed similarly to the ESCS (Economic, Social and Cultural Status) index used by the OECD Programm for International Student Assessment (see Brunello, Rocco, 2021)

We begin with specific questions drawn from the PLUS Questionnaire, from which we extract information regarding the occupation and education of both the father and mother of each individual in our sample.

Table 1A: Mother's Occupation

D105Bis. Qual è stata l'attività prevalente di sua madre nella vita?	Freq.	Percent	Cum.
A. Dirigenti	136	0,89	0,89
B. Imprenditori	270	1,76	2,65
C. Professioni intellettuali	564	3,69	6,34
D. Insegnanti	2,822	18,45	24,79
E. Professioni tecniche	267	1,75	26,53
F. Impiegati	2,935	19,19	45,72
G. Commercianti	613	4,01	49,73
H. Artigiani	236	1,54	51,27
I. Operai specializzati	325	2,12	53,39
J. Agricoltori	229	1,50	54,89
K. Operai	272	1,78	56,67
L. Professioni non qualificate	181	1,18	57,85
M. Forze armate	10	0,07	57,92
N. Casalinga	6,107	39,92	97,84
O. Non aveva occupazione	26	0,17	98,01
P. Non risponde	305	1,99	100.00
TOTAL	15,298	100.00	

Table 2A: Father's Occupation

D105Bis. Qual è stata l'attività prevalente di suo padre nella vita?	Freq.	Percent	Cum.
A. Dirigenti	835	5,46	5,46
B. Imprenditori	1,033	6,75	12,21
C. Professioni intellettuali	1,566	10,24	22,45
D. Insegnanti	931	6,09	28,54
E. Professioni tecniche	633	4,14	32,67
F. Impiegati	4,611	30,14	62,81
G. Commercianti	1,079	7,05	69,87
H. Artigiani	823	5,38	75,25
I. Operai specializzati	1,200	7,84	83,09
J. Agricoltori	457	2,99	86,08
K. Operai	967	6,32	92,40
L. Professioni non qualificate	306	2,00	94,40
M. Forze armate	431	2,82	97,22
N. Casalinga	9	0,06	97,27
O. Non aveva occupazione	27	0,18	97,45
P. Non risponde	390	2,54	100,00
TOTAL	15,298	100.00	

Table 3A: Mother's education

D105B. Mother education	Freq.	Percent	Cum.
A. Nessun titolo	161	1,08	1,08
A. Elementari	2,509	16.85	17.94
B. Medie	3,513	23.60	41.53
C. Diploma	5,951	39,97	81,51
D. Laurea	2,753	18,50	100.00
TOTAL	14,887	100.00	

Table 4A : Father's education

D105A. Father education	Freq.	Percent	Cum.
A. Nessun titolo	131	0.88	0.88
A. Elementari	1,994	13,45	14,33
B. Medie	3,474	23,43	37,77
C. Diploma	5,610	37,84	75,61
D. Laurea	3,615	24,40	100.00
TOTAL	14,824	100.00	

At this point, the construction of the famback variable consists of four steps:

1) Construction of the variables m_edu and f_edu (mother's and father's years of completed education). We associate each response to questions D105a and D105b with the corresponding number of years invested in education (for example, elementary school is associated with five years of education, while a high school diploma is associated with thirteen years of education). A STATA script illustrates the procedure.

```
* Parental education

gen f_edu = 0 if father_education==0
replace f_edu = 5 if father_education==1          ***Elementary schools last for 5 years***
replace f_edu = 8 if father_education==2          *** 5+3=8 years of completed education***
replace f_edu = 13 if father_education==3         *** 5+8+5 years of completed education***
replace f_edu = 17 if father_education==4


gen m_edu = 0 if misced==0
replace m_edu = 5 if misced==1
replace m_edu = 8 if misced==2
replace m_edu = 13 if misced==3
replace m_edu = 17 if misced==4
```

2) Construction of the *m_isei* and *f_isei* variables (mother's and father's occupational status). We associate each response to questions D105bis (father and mother) with the corresponding numerical level computed on the 2008 International Socioeconomic Index of Occupational Status (ISEI, Ganeboom and Treinman, 2003, see Brunello, Rocco, 2021).

3) We construct the variables *min_edu* and *min_isei* (minimum parental occupational and educational level) in order to compute the economic, social and cultural assortment within the household (Brunello, Rocco, 2021). The variables are defined as

- min_edu = min (f_edu; m_edu)

- min_isei = min (f_isei ; m_isei)

4) We then apply a Principal Component Analysis on the following variables*: m_edu, f_edu, m_isei, f_isei, min_edu, min_isei*. As a result of this analysis, we obtain the famback index, which is describes in Table 5A.

Table 5A: famback

```
. sum famback
```

Variable	Obs	Mean	Std. Dev.	Min	Max
famback	15,298	1.073679	2.445855	-4	6.141415

REFERENCES

ALTONJI, J.G., 1992. *The effects of High School Curriculum on Education and Labor Market Outcomes*. Cambridge: National Bureau of Economic Research.

ANGRIST, J.D., PISCHKE, J.S., 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

BONIFAZI, C., DE ROCCHI, D., PANZERI, G., 2020. *Integration of Migrants into the Labour Market in Europe*. Bingley, Emerald Publishing Limited, pp. 65-80.

BRUNELLO, G., ROCCO, L., 2017. *The effects of vocational education on adult skills, employment and wages: what can we learn from PIAAC?* SERIEs.

BRUNELLO, G., ROCCO, L., AGARWAL, L., 2021. *The Pathways to College*. Journal of Human Capital. Bonn, Institute for the Study of Labour, IZA Discussion Papers, No. 12691.

BRUNELLO, G., et al., 2023. *Do classical studies open your mind?* University of Padua.

CARDINALE, U., SINIGAGLIA, A., 2016. *Processo al Liceo Classico. Resoconto di un'azione teatrale*. (s.l): Il Mulino.

CASALEGNO, A., 18/03/2010. L'Attualità intellettuale del Classico. IL SOLE 24 ORE, 16.

D'AMICO, N., 2009. *Storia e storie della scuola italiana. Dalle origini ai giorni nostri*. (s.l.): Zanichelli.

DE PHILIPPIS, M., et al., 2022. *La crescita dell'economia italiana e il divario Nord-Sud: trend storici e prospettive alla luce dei recenti scenari demografici*. Banca D'Italia Occasional Papers No. 683.

HANUSHEK, E.A., et al., 2017. *Coping with change: International Differences in the Returns to skills*. Economics letters, 153, 15-19.

JOENSEN, J.S., NIELSEN, H.S., 2006. *Is there a casual effect of High School Math on Labour Market Outcomes?* Bonn, Institute for the Study of Labour, IZA Discussion Papers, No. 2357.

L. 11 dicembre 1969, n. 910.

L. 13 novembre 1859, n. 3725 del Regno di Sardegna.

LEVINE, P.B., ZIMMERMAN, D.J., 1995. *A comparison of the sex-type of occupational aspirations and the subsequent achievement, work and occupations*. Fèvrier: 22, 73,84.

MIUR, 2023. *Iscrizioni all'anno scolastico 2023/24, i primi dati: in aumento le domande agli Istituti tecnici, il 57,1% sceglie i Licei*.

MIUR, 2023. Allegato_1_Dati_Iscrizioni_online_as2023_24.pdf. Available on: https://www.miur.gov.it/-/iscrizioni-all-anno-scolastico-2023-2024-i-primi-dati-in-aumento-le-domande-agli-istituti-tecnici-il-57-1-sceglie-i-licei

MIUR, 2023. Allegato_2_Iscrizioni_online_Regioni_as2023_24.pdf. Available on: https://www.miur.gov.it/-/iscrizioni-all-anno-scolastico-2023-2024-i-primi-dati-in-aumento-le-domande-agli-istituti-tecnici-il-57-1-sceglie-i-licei

OLLIKAINEN, J.P., et al., 2022. Effect of Secondary Education on Cognitive and Non-Cognitive Skills. Bonn, Institute for the Study of Labour, IZA Discussion Papers, No. 15318.

ROSE, H., BETTS, J.R., 2004. *The Effect of High School Courses on Earnings*. Review of Economics and Statistics: (86: 2), 497-513.

RUBIN, D.B., 2005. *Casual Inference Using Potential Outcomes: Design, Modeling, Decisions*. Taylor & Francis, Ltd. on behalf of the American Statistical Association. Journal of the American Statistical Association, Vol. 100, No. 469, pp. 322-331.

RUBIN, H. J., RUBIN, I. S., 2005. *Qualitative Interviewing: The Art of Hearing Data (2nd ed.)*. Thousand Oaks, CA: Sage.

SILLIMAN, M., VIRTANEN, H., 2022. Labor Market Returns to Vocational Secondary Education. American Economic Journal: Applied economics, vol. 14, 1.

STOCK, J.H., WATSON, M.W., 2007. *Introduction to Econometrics*. Pearson/Addison Wesley.

ZIZZA, R., 2013. *The Gender Wage Gap in Italy*. Bank of Italy Occasional Papers No. 172.