



Exploring the impact of high schools, socioeconomic factors, and degree programs on higher education success in Italy

Cristian Usala ¹*, Isabella Sulis, Mariano Porcu

Department of Social and Political Sciences, Università degli Studi di Cagliari, Italy

ARTICLE INFO

Keywords:

School effect
Dropout
Churn
University career
MOBYSU.IT
INVALSI

ABSTRACT

This study investigates the determinants of tertiary education success in Italy, focusing on students' outcomes between the first and second years. We use population data of students enrolled between 2015 and 2019, integrating information on high school environments and degree program characteristics. This rich dataset has been exploited with a two-step approach: the first step defines indicators for high school quality and degree program difficulty; the second estimates a multinomial logit to assess the determinants of students' probability of being classified as regulars, churners, at risk of dropout, and dropouts. Data regarding the 2019 cohort have been further investigated by exploiting the additional information on students' socioeconomic backgrounds and schools' self-assessed effectiveness evaluations. Results indicate that students' high school backgrounds, socioeconomic conditions, and post-graduation prospects in terms of net wages and occupation rates of graduates in the chosen degree program significantly influence academic success and students' academic persistence. Overall, the results offer a comprehensive view of the determinants of university success, with specific patterns observed across the different student categories.

1. Introduction

Numerous studies have examined the factors influencing university students' performances due to their significance for students' future outcomes and the composition of a country's population in terms of human and social capital [1,2]. Among the elements highlighted by the literature, paramount attention has been attached to students' socioeconomic and educational background and how they interact with the characteristics of the chosen university pathway [3–5]. This paper provides an analysis of the relevance of these elements on students' career outcomes within the Italian context, where significant disparities in access to tertiary education and completion rates are influenced by students' socioeconomic and high school backgrounds [6–8].

The Italian data show that achieving a tertiary education degree is strongly associated with students' family backgrounds and previous educational experience. According to Contini et al. (2018) [3], while the tertiary education system has become more effective for most students, it remains ineffective for those from disadvantaged backgrounds and is still far from achieving the Agenda 2030 goal of increasing the country tertiary educational attainment [9]. Moreover, Busetta et al. (2023) [10] show that the likelihood of obtaining a degree for individuals from

low-educated families has remained stable among those born between 1960 and 1989, with only a slight improvement of about 6% over two decades. In contrast, no relevant improvement has been observed for families with intermediate educational levels, highlighting the stability of the so-called *education persistency* gap between generations. These differences based on students' socioeconomic conditions stem from the choice of the high school curriculum, with students from better socioeconomic backgrounds choosing to enrol in academic-oriented schools (i.e. the *Liceo* high schools) and are reflected in the high selectivity in the access to the tertiary education system [11,3,12]. Indeed, the data on students' socioeconomic backgrounds index (ESCS) measured by the National Institute for the Evaluation of Educational System (INVALSI)¹ shows that those enrolled in the university score, on average, half a standard deviation higher than those who do not enrol. Students who join tertiary education belong to families that score an ESCS index 0.30 points above the average in grade 13th. It is worth highlighting that, as illustrated in Fig. 1, the gap between enrolled and not-enrolled students remains similar across all Italian regions. However, students from the South of the country have, on average, lower values of the ESCS index than the national average.

* Corresponding author.

E-mail address: cristian.usala@unica.it (C. Usala).

¹ Students' socioeconomic background is measured by INVALSI using the Economic Social and Cultural Status Index (ESCS), which assumes a normal distribution with a national average of 0. See Section 3 for more details.

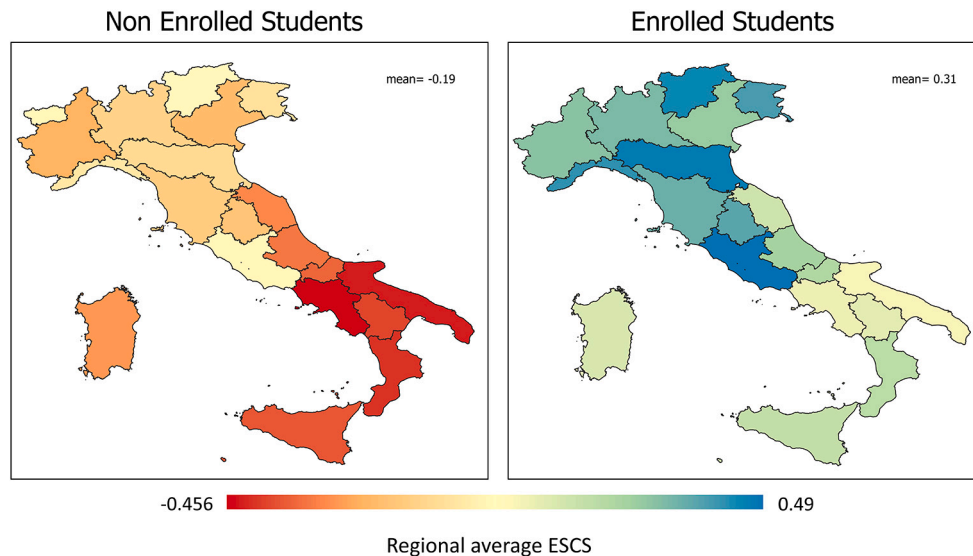


Fig. 1. Mean of the ESCS at regional level (Non enrolled = -0.19 vs Enrolled = 0.31).

In this framework, the contribution of this paper is to identify how the high school environment and degree programs' characteristics shape students' outcomes at the university. In particular, we start investigating the following research question:

RQ1 *What role do schools and degree programs play in determining students' career outcomes?*

To answer this question, we combined the administrative data on the population of students enrolled in Italian universities between 2015 and 2019 with a rich set of covariates that allows us to account for the role played by students' demographics and educational backgrounds, the characteristics of the degree programs and other confounding factors such as the local socioeconomic conditions of students' residence area and university hosting cities. The analysis is based on a two-step approach. In the first step, we define a set of indicators to measure the overall impact on students' performances in terms of CFUs of the attended high school and degree program. These indicators are obtained by estimating two linear regressions using high school and degree program fixed effects. The second step makes use of the estimated fixed effects to single out the effects on students' academic outcomes related to high schools and degree programs from those associated with students' characteristics. This is done using a multinomial logit model in which students are classified based on their first-year performance and their decision regarding whether to continue, dropout, or change university during their university journey. Then, to investigate how these elements interact with students' socioeconomic background and high school quality, we exploit the detailed information available for the cohort of students enrolled in 2019, which includes information on students' and peers' socioeconomic statuses and indicators related to students' outcomes and schools' processes obtained from schools' self-assessment evaluations carried out within the National System of Evaluation. This approach allows us to address two additional research questions:

RQ2 *To what extent can high school quality indicators be related to students' academic outcomes at the university?*

RQ3 *To what extent do students and peers' socioeconomic status influence their career outcomes?*

Results show that high school background, students' socioeconomic conditions, and the characteristics of universities and degree programs significantly shape students' academic success. Students from academic schools (the *Liceo* high schools) have better outcomes, while those from

technical and vocational institutes face higher dropout risks. Students with higher final high school marks and those who attended high-performing high schools are also less likely to experience adverse outcomes, such as dropping out of the university system or changing universities during one's academic career. The same is true for degree programs where students accumulate a higher average number of credits and graduates have better outcomes after graduation regarding wage prospects and occupation as measured by the AlmaLaurea surveys (see Section 3).

The paper is structured as follows. Section 2 outlines the theoretical framework underpinning this research. In Section 3, we describe the data sources and data selection process and provide some descriptive statistics. Section 4 details the methodological approach adopted to achieve the primary objectives of the study. Section 5 presents and discusses the main findings, while Section 6 contains the concluding remarks.

2. Theoretical framework

In recent decades, national and international literature has emphasized that graduating from university is merely the end of a prolonged educational journey, heavily influenced by various factors, actors, and decisions. Studies reveal that the highest educational level achieved results from decisions made by both students and their families, which in turn are influenced by the surrounding environment and significant events experienced during earlier educational stages. In this framework, several factors affect students' university careers such as students' characteristics, family background, relational factors related to academic integration, tertiary education system attributes, and labour market expectations [1].

University students' decision-making has been primarily analysed from two perspectives. The Economic Perspective, primarily rooted in Human Capital Theory [13], highlights the importance of informational factors in shaping students' academic careers. It focuses on the informational factors and the disparity between the expected returns from education and the associated costs. Extensions of the human capital model take into account that students periodically update their expectations based on their experiences throughout the learning process, highlighting that the expected present value of obtaining a tertiary degree is revised to include new information on monetary and non-monetary benefits and costs associated with investing in higher education. Among these elements that influence students' expected returns, labour market expectations and the socioeconomic characteristics of students' residence areas and universities' hosting territories play a significant role by influencing the costs and benefits associated with pursuing higher education.

However, the direction of this influence is not always clear since it depends on the interaction between geographical factors, students' socioeconomic background and educational choices [14,15]. For example, while Contini et al. (2018) [3] show that higher unemployment rates are linked to lower graduation rates and reduced continuation of studies in Italy, Di Pietro (2006) [16] notes that this relationship shifts when accounting for regional labour market differences, with Italian students less likely to quit their studies in regions with lower employment rates [16,7]. Moreover, these expectations are linked to the job market opportunities associated with the specific disciplinary field (e.g., STEM and no STEM) or degree program chosen [15,16,7].

However, this perspective is limited since it underestimates the effect of non-monetary benefits and assumes that students have complete information and the ability to make fully rational decisions. Indeed, the Relational Perspective, represented by Tinto's Model [17] and the related literature, emphasises the significance of relational factors and non-monetary benefits, which are strictly linked to the quality of social interaction. It stresses the importance of social and academic integration, as well as the role of a sense of belonging and support from peers and faculty [18–20]. This literature highlights that students who are engaged in university activities and develop relationships with peers and professors tend to feel a stronger sense of belonging to the institution, which translates into greater commitment. Relational factors that produce non-material benefits and reduce non-monetary costs include the intensity of relationships with faculty, academic performance discussions, study group formation, orientation days, the implementation of academic policies to identify students at risk of dropping out, and the development of targeted support policies (e.g., mentoring, tutoring). Involvement in extracurricular activities, often facilitated by student associations, is also beneficial.

Therefore, while monetary benefits are primarily linked to labour market conditions, the student's socioeconomic background, and the chosen field of study, non-monetary benefits largely depend on the student's expectations regarding the academic environment, relationships with peers and faculty, and the support policies in place. Building on this framework, we define an empirical strategy that allows us to identify the determinants of students' academic success during the first year of career.

3. Data

The analysis of the elements that affect students' academic success relies on the combination of several sources.

Information on students' university careers and characteristics come by the MOBYSU.IT database [21], which includes the administrative data on the cohorts of students enrolled for the first time in an Italian bachelor's degree program between 2015 and 2019. Starting from the population of 1,186,040 students, we define a set of eligibility criteria that allows us to explore the role played by students' high school environments and degree programs' characteristics.² First, since we need information on students' high school background and geographical area of residence, we excluded the students with less than 17 years at the moment of enrolment (395), those for which we do not observe the city of residence (17,251) or the attended high school (78,233), and the schools that have only one graduate enrolled at the university (2,174). Second, we excluded 37,946 students enrolled in online universities. This exclusion is because we cannot identify universities' locations, which are crucial to understanding the role of the socioeconomic conditions of universities' hosting areas. Third, we do not consider the 97,908 students who enrolled in degree programs that are not covered by the AlmaLau-

rea survey on graduates' outcomes after graduation. As explained below, this data allows us to analyse the relationship between post-graduation outcomes in terms of wages, occupation rates, and master enrolment at the degree program level. The final data refers to 952,133 first-year students.

For each student, MOBYSU.IT includes the information on students' chosen university and degree program, as well as the number of formative credits (CFU) earned during their first career year.³ This information allows us to identify how students' careers proceed in the transition between their first and second years based on both students' choices and performances. In particular, each student is classified into four categories: dropout, at risk of dropout, churn, and regular. Students are classified as dropouts if they are not enrolled in an Italian university at the beginning of their second year. This classification has two limitations: first, we do not account for students who may have enrolled in universities abroad; second, we cannot identify students who remain formally enrolled but are not actively taking exams or have otherwise ceased their university activities (i.e., implicit dropouts). Although we cannot account for the first limitation, we can address the second one using the category of students at risk of dropout. These students are those who have a low performance with respect to their colleagues in terms of formative credits gained but have decided to stay enrolled at the university. In particular, we defined at-risk students as those who earned less than 25 CFUs during the first year of their degree program. Indeed, as shown in Fig. 2, the proportion of dropouts in the data is higher than the one of non-dropouts when we consider students with less than 28 CFUs.⁴ Finally, churning students are those who decided to change the university between the first and second year, while regulars are the residual category.

Besides the data on students' careers, MOBYSU.IT provides information on their socio-demographic characteristics, such as sex, age, and city of residence. Moreover, it helps us draw a detailed profile of students' educational backgrounds using the data regarding their attended high school, final high school mark, and high school curricula.⁵ The combination of information on students' high school curricula and macro area of residence helps us to partially account for the socioeconomic conditions of students. Indeed, as explained in Sections 3.1 and 1, students in the Centre-North of the country and those who graduated from the *Liceo* high schools come, on average, from better socioeconomic backgrounds in terms of cultural capital and parents' educational and professional statuses. Table 1 reports the descriptive statistics regarding students' characteristics for each category.

From the Table 1, we can see that the percentage of regular students is the largest category at 67.3%, while churning represent only 4.3% of the sample. Moreover, 17.2% of students are identified as at risk of dropout, and 11.2% of students drop out by their second year. The data on students' performances show how regular students outperform all the other groups by earning, on average, 52 CFUs against an average of 39 CFUs. The gap in performance between regular students and churning is relevant, indicating that students who change universities may be incentivised to do so by their low performance in the chosen university. Females are more represented among regular students (58.14%) with respect to the other categories, especially considering the dropouts. High school performance, as indicated by the final high school mark, is pos-

² The degree programs are based on the Italian classification of "Classi di Laurea" that groups the specific courses according to their disciplinary contents. See Table A.4 in the Appendix for the classification of degree programs according to the ISCED 2-digit classification and the average number of students observed in each program between 2010 and 2019.

³ The CFUs (*Crediti Formativi Universitari*) are used in the Italian tertiary education system to measure the workload associated with each exam or activity (similar to the European Credit Transfer and Accumulation System and Grade (ECTS)). Usually, one CFU indicates 25 hours of study with differences across disciplinary fields.

⁴ We have estimated the models using other thresholds around 30 CFUs obtaining similar results. The results are available from the authors upon request.

⁵ We have grouped the high school curricula into five categories. The *Liceo Classico* and *Liceo Scientifico* high schools provide a more academic-oriented program, and Technical and Vocational institutes give a more professional labour-oriented program. Other institutes represent a residual category.

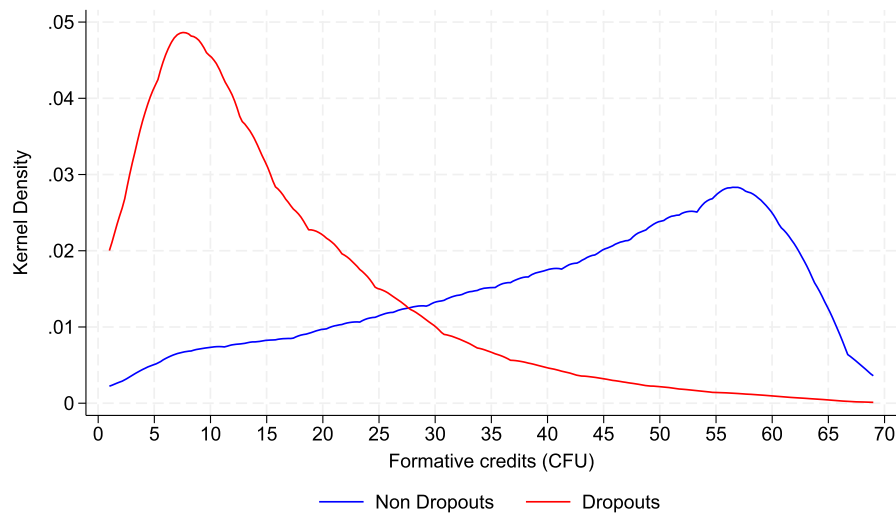


Fig. 2. Kernel densities of CFUs for dropouts and non-dropouts. Notes: The figure reports the kernel distributions of CFUs for dropouts (students who are not enrolled in an Italian university in the second year) and non-dropouts (students who are enrolled in an Italian university in the second year). The plot shows the information on students who have obtained at least 1 CFU and less than 70 CFUs.

Table 1

Descriptive statistics on students' characteristics by status.

	Regular	Churn	Risk of dropout	Dropout	All
Enrolled	67.3%	4.3%	17.2%	11.2%	952,133
Students' characteristics:					
Average CFU	52.2	17.2	12.4	7.2	38.8
Females (%)	58.1	56.9	48.6	44.2	54.9
High school career:					
Average H.S. final mark [60;101]	81.4	79.2	74.9	74.0	79.4
Late diploma (%)	11.0	10.7	22.2	29.1	14.9
High school type (by col):					
<i>Liceo Classico</i> (%)	10.6	14.8	9.4	5.5	10.0
<i>Liceo Scientifico</i> (%)	38.4	42.1	33.9	21.3	35.9
Other Institutes (%)	23.8	21.3	20.1	22.2	22.9
Technical Institutes (%)	21.1	16.2	26.8	35.5	23.5
Vocational Institutes (%)	6.1	5.5	9.9	15.5	7.8

Notes: The table shows the descriptive statistics regarding students' characteristics. It shows the average value for continuous variables (e.g., CFU) and the percentage of students in each category for categorical variables (e.g., Females). Columns 1-4 show this information with respect to the students included in each category (regulars, churners, at risk of dropout, dropout). Column 5 reports the descriptive statistics considering all students. H.S. indicate the High School. See Table A.2 in the Appendix for the definitions and descriptive statistics of each variable.

itively correlated with university outcomes. Regular students have the highest final mark (81), while dropouts show the lowest (74). The share of students who delayed their high school graduation is also notably higher among dropouts (29.2%) and at-risk students (22.2%) compared to regular and churner students. Finally, the distribution of high school types among the student population varies by status. Students from *Liceo* high schools and other institutes are overrepresented among regular and churn categories, while dropouts and students at risk of dropouts are more common in technical and vocational institutes. This underscores the impact of high school background on university trajectories, with students from *Liceo* high schools tending to achieve better outcomes and to re-orient themselves more often than the other categories. This element, as explained in Section 3.1, can be related to the fact these students come, on average, from families with higher socioeconomic backgrounds, which provide greater flexibility and resources to explore preferred degree programs. Overall, the descriptive statistics suggest the importance of high school background and students' educational experiences in shaping their higher education success.

To analyse the role played by universities and degree program characteristics, we integrated the MOBYSU.IT data with other relevant information. Specifically, we have included information on the average revenue per student and the number of scholars and non-scholars at the university level from the USTAT database.⁶ In particular, the average revenue represents the total revenue from student enrolment divided by the number of enrolled students as measured by USTAT in 2018. Moreover, to account for the opportunities and expectations related to each degree program, we also included the information provided by AlmaLaurea for each degree program. Indeed, AlmaLaurea is an inter-university consortium that provides comprehensive surveys on the Italian university system. These surveys offer a detailed system picture through annual surveys on graduates at each level (bachelor's, master's, etc.) and information on graduates' satisfaction, profiles, and employment opportunities. In this work, we focus on the data regarding the occupation rate one year after graduation, the share of graduate students who decide to enrol in a master's program, and the net wage of graduates one year after graduation.⁷ Finally, to account for the confounding factors related to territories' socioeconomic conditions we included in the analysis the data regarding the provincial unemployment rate and the municipal number of units (firms or branches of firms) in students' residence area from National Statistical Institute (ISTAT) and the information on the municipal distribution of taxable income per capita in universities' hosting cities from the Italian Ministry of Economy and Finance. All the data regarding territories' socioeconomic characteristics refer to year 2018, one year before the one in which students are observed, to mitigate issues related to simultaneity.

3.1. A focus on the 2019 cohort

To provide a deeper investigation into the role of high school students' and peers' characteristics on higher education success, we analysed the data separately for the cohort of students enrolled in the

⁶ USTAT is an open data portal based on the administrative data collected by the Italian Ministry of University and Research which contains several information on the characteristics of Italian universities. See <https://ustat.mur.gov.it/> for more information.

⁷ On average, 83% of graduates completed the survey in our time frame. However, we do not have information on degree programs with less than five graduates in the considered year (e.g., defence and security degree programs). Moreover, several universities do not participate in the survey: Bocconi, Cattolica, Polytechnic of Milan, Dante Alighieri, Mercatorum, Link Campus, Europa, LUISS, Saint Camillus, and Humanitas University.

2019/2020 academic year. Indeed, for this cohort, MOBYSU.IT includes a specific section with additional details on students' socioeconomic conditions provided by the National Institute for the Evaluation of the Education and Training System (INVALSI). In particular, INVALSI provides the index of students' family socioeconomic and cultural status (ESCS). The ESCS measures students' socioeconomic and cultural backgrounds based on parents' educational and professional statuses and whether certain items are present in the student's home, such as having a quiet place to study, a desk for doing homework, or a well-stocked bookshelf. The data on the ESCS index is available on three levels: individual, class, and high school. Therefore, the ESCS index can account for peers' and students' socioeconomic characteristics, also considering those not enrolled in any Italian university.⁸

To include this crucial information in the analysis, we excluded from the data 46,967 students for whom we did not observe the information on the ESCS index. These missing values are present because the INVALSI data was collected only from students who participated in the 2018/2019 survey. Therefore, we do not have information on those who graduated from high school before 2019 but enrolled at the university in 2019. The data on the ESCS index shows relevant differences in students' socioeconomic backgrounds, which vary by geographical region and high school curriculum. Considering an average ESCS value of 0.26, we find that students from the Centre-North have an average of 0.29, while those from the South and Islands score a much lower average of 0.13. Regarding high school background, we observe a relevant difference between academic tracks (the *Liceo* high school), with an average index of 0.46, and technical and vocational institutes, which have a negative average value equal to -0.05.

To explore the role of high school quality in shaping students' performance at universities, we integrated data regarding the high school self-assessment exercises carried out in 2016 and 2017 within the National System of Evaluation. This information can be used to contextualise schools with respect to their outcomes and processes. In particular, we have data on 11 indicators, 4 related to evaluating schools' outcomes and 7 to assessing their processes.⁹ These indicators range from 1 (very critical situation) to 7 (excellent situation) based on analytical guidelines provided by the system in the so-called "Assessment Rubric". These indicators are summarised through a Principal Component Analysis (PCA) in two indexes, one that measures high schools' outcomes and the other that is related to high schools' processes.¹⁰ This data is not available for 7,429 students who have obtained their high school diploma from a school that did not participate in this self-evaluation exercise. Therefore, the final data of the second part of the analysis includes 115,609 students who represent 68% of the population of high school leavers in 2019 who have enrolled in an Italian university. Table A.3 in the Appendix compares students included in the estimation with those excluded from the analysis. The main differences are the proportion of students who graduated late from high school, which is higher among included students, and the distribution of students across high school types, with excluded students having a lower proportion in vocational institutes. These differences suggest that results based on completed cases may underestimate the proportion of regular students.

⁸ The data at the class and high school level is obtained considering all the students participating in the INVALSI survey at grade 13. Therefore, it also includes students who decided not to enrol at the university and those who enrolled abroad.

⁹ The outcome indicators refer to these four dimensions: proficiency, standardised test scores, key and citizenship skills, and long-term results. The dimensions related to high school processes are: learning environment, inclusion and differentiation, continuity and orientation, strategic orientation and organisation of the school, human resources development and enhancement, integration with the territory, and relationships with families.

¹⁰ The results of the PCA analysis are available upon request. For both indexes, we have retained the first components resulting from the PCA, which explains around 61% of the variance in both cases.

However, due to the lack of data on students' socio-economic background and high school quality indicators, the analysis for the 2019 cohort relies solely on completed cases. It is important to note that the analysis of all other cohorts is not affected by this issue regarding missing data. See Appendix Table A.1 for more information on students' characteristics and high schools' self-evaluation results.

4. Methodological approach

To investigate the factors determining higher education success in Italy, we employ a two-step procedure to single out the effects related to high schools' effectiveness and degree programs' "easiness" from those associated with students' and degree programs' characteristics.

First step

In the first step, we model the number of CFUs earned by students during their first year of university, taking into account the clustering of students in high schools and degree programs. Specifically, we estimate two fixed effects models: one including a set of high school fixed effects, and the other considering a set of degree programs fixed effects. Formally, the CFU earned by student i who graduated from high school h and enrolled in degree program d can be modelled using two equations:

$$CFU_i = \alpha + v_h + u_i \quad \text{with } h = 1, \dots, 18363 \quad (1)$$

where CFU_i is the number of CFUs earned by student i during their first year of enrolment, v_h is the high school fixed effects, i.e., a set of 18,363 dummy variables that identify the observed high schools, and u_i is the unobserved error term. In this case, v_h captures all the elements that are specific to the high school h and that do not vary over time. Therefore, since the higher the estimated fixed effects, the higher the number of CFUs earned by the students who attended the high school, the estimated \hat{v}_h can be interpreted as a proxy for the effectiveness of the high school in upskilling students for university-level education. We exploit this in the second step to isolate the effects related to the high school from those associated with the other elements. Moreover, this approach allows to have a proxy for the high school's effectiveness without including all fixed effects in the second step equation.

$$CFU_i = \alpha + g_d + u_i \quad \text{with } d = 1, \dots, 45 \quad (2)$$

Eq. (2) differs from (1) as it contains the set of 45 degree programs fixed effects g_d . In this case, the estimated \hat{g}_d measures the difficulty level of the chosen degree program. In particular, these indicators can be interpreted as proxies for the degree program's "easiness" since the higher the estimated value, the higher the number of CFUs earned on average by the students enrolled. Therefore, these indicators help us account for all the elements specific to a degree program that positively affect students' performances. Moreover, since we use the data at the degree program level, this strategy helps account for the difference between the observed disciplinary fields.¹¹

Second step

The second step uses the fixed effects estimated according to Eq. (1) and (2) as predictors in a multinomial logistic regression model. Specifically, we model the probability that a student i is observed in status j , where j can be one of the three categories defined in Section 3: churn, at risk of dropout, dropout. The category regular (i.e., $j = 0$) is considered as our baseline. The multinomial logistic function is:

¹¹ The results of the two fixed effects models are available from the authors upon request.

$$\log\left(\frac{\pi_{ij}}{\pi_{i0}}\right) = \alpha_j + \mathbf{X}'\boldsymbol{\beta}_j + \gamma_{1j}\hat{\nu}_h + \gamma_{2j}\hat{g}_d \quad \text{with } j = 1, 2, 3 \quad (3)$$

where π_{ij} is the probability of student i being in status j , and π_{i0} is the baseline probability. Therefore, for each status j , we estimate one equation by identifying the constant α_j , the vector of parameters $\boldsymbol{\beta}_j$ that informs on the effects of the covariates on students' probability, and the parameters γ_{1j} and γ_{2j} which estimate the effect associated with the high school and degree program fixed effects, respectively. \mathbf{X} indicates the set of covariates, which includes students' socio-demographic characteristics, the variable related to the labour market conditions and expectations, and the university characteristics. In particular, student socio-demographic characteristics are the sex, the variables related to the high school background, and a set of dummy variables indicating the macro area of residence. The variables related to the academic environment can be classified in two categories: university characteristics and post-graduation characteristics related to the degree program. The first group includes: the number of scholars, the number of non-scholar staff, and the revenues per student at the university level. The second group includes the information on graduates' outcome as measured by the AlmaLaurea survey regarding graduates' net wage, occupation rate, and enrolment rate at the degree program level. Finally, to account for the local socioeconomic characteristics we included the unemployment rate and the number of firms in students' residence area, and the average taxable income in universities hosting cities. This wide set of covariates helps us recover various information on the determinants of students' success in higher education while considering the influence exerted by the high schools and the degree programs.¹²

Moreover, to account for the two-step nature of our procedure and to ease the interpretation of the results, we computed the standard errors using 200 bootstrap replications considering the clustering of students at the university level, and we re-scaled all the continuous covariates in z-scores.

Finally, as explained in Section 3.1, we complete our analysis by focusing on the cohort of students enrolled in 2019, for which more detailed information on students' socioeconomic backgrounds and high school characteristics is available. This allows for a more detailed analysis of how the joint effects of high schools, peers, and students' socioeconomic status influence students' outcomes. This is done by estimating the second step, including the information on students' and classes' ESCS index and high schools' self-evaluation scores. Moreover, to better account for the role of high schools and degree programs, we included in the model the fixed effect estimates $\hat{\nu}_h$ and \hat{g}_d , obtained in the first step using the complete set of available observations. This approach ensures that these indicators are not exclusively based on data from the 2019 cohort but reflect information from students enrolled between 2015 and 2019, providing a more comprehensive estimate of the high school quality and degree program difficulty.

5. Results

Table 2 shows the results of the multinomial logit model estimated in the second step. The estimated results provide a comprehensive picture of the factors influencing students' academic success in the Italian tertiary education system.

With respect to students' characteristics, we note that females are significantly less likely to dropout than males. However, they show a higher probability of being classified as churning or at risk of dropout, suggesting that while females are generally more likely to continue their studies, they still face challenges maintaining a regular academic career.

Geographical differences further highlight disparities in student outcomes. Students from the South exhibit higher risks of churn with respect to students in the Centre, while those from the North East and

North West are less likely to be at risk of dropout and dropout. These results suggest that, in general, students from the Centre (the baseline) are more likely to be classified with a negative outcome, even though they show higher average values of the ESCS index.

High school background strongly influences student outcomes. Students from academic high schools oriented towards maths and sciences (the *Liceo Scientifico*) show the lowest probabilities of experiencing an adverse career event, with negative and significant coefficients across all categories. Instead, students from academic high schools oriented towards humanities and classical studies (the *Liceo Classico* – the reference group) have a higher probability of being classified as churning or at risk of dropout. This result shows that students from high school with an average better socioeconomic background have a lower likelihood of dropping out. However, the results regarding churning and at-risk students show, on the one hand, that those who graduated from this type of high school may have more difficulty finding the degree program that better fits their capabilities and tend to change degree programs more often. On the other hand, since *Liceo* high schools are academically oriented, these students may prefer to stay at the university even though they are at risk of dropping out rather than leaving the higher education system to enter the job market without a tertiary degree. Accordingly, students from vocational and technical institutes face a significantly higher likelihood of dropout than students from other categories. The high school final mark is a strong predictor of student success. Higher grades are consistently associated with a lower risk of experiencing an adverse event: students who perform well in high school are likelier to maintain a regular path in university. Finally, the results suggest that students with an irregular high school trajectory (i.e., students who completed high school after the age of 19) continue to face challenges in university as they have significantly higher probabilities of being at risk of dropping out or dropping out.

These results align with other studies carried out in Italy [3,12,22] and hold even controlling for the high school effectiveness as measured by the estimated fixed effects. With respect to this element, our findings show that students who attended high schools with higher effectiveness (i.e., higher average CFUs) have lower probabilities to experience an adverse adverse outcome, especially with respect to the categories at risk of dropout and dropouts. These results confirm the role of high schools in shaping future academic outcomes and confirm that the implemented analytical strategy can identify the influence related to high schools' effectiveness.

The results related to the average outcomes of graduated students measured by the AlmaLaurea surveys indicate that students' expectations also play a pivotal role. Indeed, students in degree programs with higher wage expectations have a lower probability of churning or dropping out. In contrast, higher occupation rates lower the likelihood of dropout and increase the probability of staying at the university even when the number of earned credits is lower than 25. These results indicate that students, on average, prefer degree programs where students reach better outcomes after graduation, such as higher wages or occupation rates. Looking at the variables related to areas' socioeconomic conditions (e.g., taxable income and unemployment rates), we can see that these characteristics do not seem to play a relevant role in shaping students' outcomes, especially if compared to the degree program level indicators.

Finally, the results on universities' characteristics show an interesting result: the number of non-scholars per student negatively affects the probability of being classified at risk of dropout or dropout. In contrast, the number of scholars has a positive effect. This result indicates that universities' service quality, related to the administrative staff, is an essential determinant of student outcomes.¹³ Finally, the coefficients associated with the estimated degree program fixed effects show that students in degree programs with a higher average number of credits

¹² See Table A.2 in the Appendix for covariates' definitions and descriptive statistics.

¹³ This result is in line with the ones found in [23,24].

Table 2
Two-steps estimation results.

	Churn		Risk of dropout		Dropout	
	β	se	β	se	β	se
Constant	-2.524***	0.107	-1.494***	0.088	-2.277***	0.074
Male	-		-		-	
Female	0.170***	0.034	0.057***	0.022	-0.235***	0.026
Centre	-		-		-	
Islands	-0.016	0.233	-0.031	0.111	-0.007	0.098
South	0.277*	0.151	-0.118	0.105	-0.040	0.129
North East	-0.164	0.184	-0.282***	0.106	-0.144	0.090
North West	-0.122	0.198	-0.105	0.093	-0.172**	0.074
High School final mark	-0.345***	0.025	-0.716***	0.017	-0.743***	0.021
Liceo Classico	-		-		-	
Liceo Scientifico	-0.428***	0.049	-0.353***	0.039	-0.288***	0.049
Other institutes	-0.240***	0.064	0.145***	0.050	0.655***	0.073
Technical institutes	-0.618***	0.143	0.099	0.152	0.742***	0.184
Vocational institutes	-0.380**	0.179	0.418**	0.212	1.160***	0.247
Regular in H.S.	-		-		-	
Non-Regular in H.S.	-0.000	0.032	0.501***	0.036	0.610***	0.043
Almaurea: In university	0.268***	0.058	0.336***	0.037	0.084***	0.032
Almaurea: Occupation	0.026	0.074	0.042	0.038	-0.115***	0.038
Almaurea: Wage	-0.089***	0.025	-0.015	0.018	-0.047**	0.019
Non-Scholars per student	-0.072	0.077	-0.134*	0.069	-0.109*	0.062
Scholars per student	0.025	0.106	0.189***	0.073	0.128*	0.068
Revenues per student	-0.019	0.044	-0.055*	0.032	-0.051*	0.029
Stud. residence: Firms	-0.001	0.028	0.017	0.024	-0.003	0.024
Stud. res.: Unemployment	0.087	0.083	0.190***	0.057	0.139**	0.063
Univ. location: Income	0.100	0.073	0.003	0.044	0.010	0.047
High School F.E. on CFU	-0.168	0.113	-0.266*	0.153	-0.424***	0.159
Degree prog F.E. on CFU	-0.343***	0.081	-0.427***	0.049	-0.261***	0.054
Bootstrap Reps	200					
Log Likelihood	-812893.3					
Observations	952059					
Pseudo R2	0.10					

Note: The table reports the coefficients and standard errors of the multinomial logit model estimated according to the two step procedure explained in Section 4. The first step estimates, via a fixed effects regression, the variables and that are related, respectively, to how high schools and degree programs affect students' CFU accumulation at the university. In the second step we use these indicators in a multinomial logit framework to understand their role in shaping students probability to experience an adverse outcome (Churn, Risk of Dropout, Dropout). H.S. and F.E. indicates, respectively, the High Schools and the Fixed Effects.

have a lower probability of experiencing a negative outcome. Although this result was expected, it calls into question the importance of the elements that may concur in determining degree program easiness, which may be related not only to the difficulty of the disciplinary subjects but also to other policies and services that may improve students' satisfaction and increase their opportunities to follow a regular path at the university.

Table 3 shows the 2019/2020 cohort results for which we introduced additional variables related to students' socioeconomic conditions (ESCS) and high school self-evaluation scores. Therefore, this analysis explains how students' socioeconomic backgrounds and high school characteristics influence their academic trajectories.

Including ESCS variables provides valuable insights into the influence of students' socioeconomic conditions. Specifically, the ESCS at the class level, which proxies for peers' socioeconomic backgrounds, has a significant and negative impact on the risk of dropout and dropout, suggesting that students from classes with peers with higher socioeconomic status are more likely to follow a regular path at the university. This relationship does not hold for churn, where the coefficient is smaller and not significant. This suggests that while socioeconomic advantages at the class level protect against dropout risks, they do not have a notable influence on churn, which, in turn, may be more influenced by the conditions that students experience in the chosen degree program. Additionally, the effect of class-level ESCS is similar to the one attached to high school fixed effects, indicating that while the high school background has a relevant influence on how students are prepared for their academic career, degree programs' and universities' characteris-

tics play a much relevant role in the choice of the university where to enrol.

The results on students' socioeconomic backgrounds show a different picture. Indeed, the individual ESCS positively correlates with the probability of churn, indicating that students from higher socioeconomic backgrounds may be more able to support the costs related to switching programs or universities during their academic careers. Conversely, it slightly reduces the probabilities of being classified at risk of dropout or dropout, in line with the previous findings that students from better socioeconomic backgrounds tend to have more stable academic paths. Therefore, while the socioeconomic environment during high school has a more substantial effect on students' performances, the individual one seems more important in helping students to satisfy their preferences in choosing their degree program, even after enrolment.¹⁴

With respect to high school self-evaluation scores, the results show that the quality of high school outcomes significantly decreases the likelihood of all adverse outcomes. This suggests that schools that rate themselves highly in terms of student outcomes tend to produce students with more regular academic careers. Indeed, the outcome indicators measure students' results in both the short term (during high school) and the long term (during university). Therefore, this result suggests that the self-evaluation of high schools correctly measures their ability to provide

¹⁴ To explore the heterogeneity of the effects related to the ESCS, we interacted the index with students' geographical areas and with indicators for high schools and degree programs. However, no significant differences were observed among these dimensions. The results are available from the authors upon request.

Table 3
Two-steps estimation results (2019 cohort).

	Churn		Risk of dropout		Dropout	
	β	se	β	se	β	se
Constant	-2.362***	0.116	-1.370***	0.108	-2.311***	0.099
ESCS: Class	-0.042	0.027	-0.114***	0.013	-0.146***	0.017
ESCS: Student	0.148***	0.021	-0.032***	0.010	-0.077***	0.016
Male	-	-	-	-	-	-
Female	0.212***	0.048	-0.024	0.021	-0.387***	0.026
Centre	-	-	-	-	-	-
Islands	-0.172	0.249	0.049	0.129	-0.095	0.110
South	0.115	0.154	-0.142	0.120	-0.098	0.131
North East	-0.319	0.195	-0.522***	0.108	-0.189**	0.088
North West	-0.097	0.200	-0.241**	0.113	-0.094	0.097
High School final mark	-0.360***	0.034	-0.719***	0.020	-0.766***	0.026
<i>Liceo Classico</i>	-	-	-	-	-	-
<i>Liceo Scientifico</i>	-0.553***	0.060	-0.516***	0.047	-0.410***	0.060
Other institutes	-0.373***	0.068	-0.094*	0.050	0.365***	0.057
Technical institutes	-0.631***	0.092	-0.155	0.099	0.476***	0.111
Vocational institutes	-0.441***	0.114	0.154	0.128	0.890***	0.145
Regular in H.S.	-	-	-	-	-	-
Non-Regular in H.S.	0.025	0.056	0.548***	0.029	0.543***	0.035
HS self eval: Outcomes	-0.059**	0.029	-0.117***	0.026	-0.141***	0.032
HS self eval: Processes	0.045*	0.025	0.094***	0.017	0.086***	0.022
Almaurea: In university	0.199**	0.080	0.423***	0.049	0.022	0.040
Almaurea: Occupation	0.003	0.092	0.212***	0.056	-0.119***	0.042
Almaurea: Wage	-0.131***	0.034	-0.062***	0.023	-0.091***	0.023
Non-Scholars per student	-0.074	0.101	-0.136*	0.074	-0.131**	0.062
Scholars per student	0.057	0.128	0.250***	0.075	0.203***	0.059
Revenues per student	-0.130	0.079	-0.045	0.037	-0.032	0.027
Stud. residence: Firms	-0.044	0.030	0.008	0.028	-0.014	0.023
Stud. res.: Unemployment	0.097	0.078	0.146***	0.051	0.157**	0.066
Univ. location: Income	0.060	0.084	-0.011	0.049	0.007	0.046
High School F.E. on CFU	-0.091**	0.042	-0.194**	0.096	-0.362***	0.102
Degree prog F.E. on CFU	-0.334***	0.035	-0.451***	0.026	-0.302***	0.025
Bootstrap Reps	200	-	-	-	-	-
Log Likelihood	-90801.3	-	-	-	-	-
Observations	115310	-	-	-	-	-
Pseudo R2	0.10	-	-	-	-	-

Note: Note: The table reports the coefficients and standard errors of the multinomial logit model estimated according to the two step procedure explained in Section 4. The first step estimates, via a fixed effects regression, the variables and that are related, respectively, to how high schools and degree programs affect students' CFU accumulation at the university. In the second step we use these indicators in a multinomial logit framework to understand their role in shaping students probability to experience an adverse outcome (Churn, Risk of Dropout, Dropout). H.S. and F.E. indicates, respectively, the High Schools and the Fixed Effects.

students with good preparation for academic careers. In contrast, the results regarding high school processes show a positive relationship with all adverse outcomes, indicating a possible mismatch between school process assessments and actual students' performances. These results indicate that the self-evaluation of processes does not necessarily correlate with adequate academic support or better student outcomes, highlighting a potential gap in the alignment between school processes and the actual preparedness of students for higher education.

6. Conclusions

In this paper, we analysed the factors influencing students' outcomes at university, focusing on how high schools and degree programs may mitigate or exacerbate inequalities linked to students' socioeconomic backgrounds. Using a two-step approach, we first defined indicators to capture high school effectiveness and degree program challenges. We then examined how these factors shaped their academic trajectories controlling for potential confounding factors. For the 2019 cohort, we further integrated detailed data on students' and peers' socioeconomic statuses and high school self-evaluation scores to assess the role of socioeconomic background and school quality in shaping university outcomes.

Our findings show that high school background, socioeconomic factors, and students' expectations strongly influence university outcomes, even controlling for potential confounding factors related to the local

socioeconomic conditions. Students from academically focused schools, primarily academic high school tracks oriented towards maths and sciences, tend to perform better, while those from technical and vocational schools are more likely to drop out. Higher high school final marks and attending more effective high schools reduce the chances of adverse outcomes. Labour market expectations associated with the attended degree program are crucial in encouraging students to stay in their degree programs.

These results allow us to address our research questions and explore how the different factors included in the analysis impact students' academic success in higher education.

Concerning the RQ1 "What role do schools and degree programs play in determining students' career outcomes?" the results indicate that high schools and degree programs strongly influence students' career outcomes. Students from academically oriented high schools show better outcomes, while vocational and technical institutes are associated with higher dropout rates. Moreover, while high school effectiveness, as measured by the average number of CFUs, reduces students' probability of dropout, it does not significantly affect churn decisions. Indeed, students tend to be more affected by degree program characteristics when deciding whether to change university.

Regarding the RQ2 "To what extent can high school quality indicators be related to students' academic outcomes at the university?", our analysis of the data on the 2019 cohort shows that high school self-assessed quality regarding students' outcomes strongly correlates with university

success. Students from high schools with better self-assessed evaluations are more likely to follow regular academic paths, reinforcing the role of secondary education quality in preparing students for higher education. However, the same is not valid for school self-assessed quality regarding school processes and policies. Indeed, students from high school with higher scores have significantly higher dropout rates, although the magnitude of the estimated coefficients is lower than the one associated with students' outcome indicators.

Finally, for the RQ3 "To what extent do students and peers' socioeconomic status influence their career outcomes?", the results show that students and peers' socioeconomic status positively influences their academic trajectories. In particular, while students' status significantly affects all the outcomes considered, peers' socioeconomic status has a significant effect only when considering the dropout risk and the dropout rate. These results indicate that students from higher socioeconomic backgrounds, particularly those with higher individual ESCS scores, are better equipped for the academic challenges and also have the necessary resources to decide to change university if necessary. In contrast, students from less advantaged backgrounds face greater dropout risks, even when attending high-performing high schools. This result aligns with the findings of Engberg & Wolniak [25], who in shedding light on the effect of social context on individual behaviour in educational framework, identify peers' networks, along with students' social and individual capital and college-linking activities within the school feeder network, as the primary factors influencing college performance. Moreover, in further research, the authors [26] provide evidence that the socioeconomic context of a high school acts as a moderating factor between individual and school-level resources in influencing students' propensity to enrol in college. This effect helps compensate for the imbalance in socioeconomic resources among secondary schools at both the individual and school levels. Their work provides evidence that between school-differences have a greater influence that within school differences in influencing for instance college enrolment. However, the same authors [27], in analysing the association between high school environment and college performance, highlight that the most advantaged students benefit the most from the high school context, infrastructure, and quality of teaching, recognizing that family socioeconomic resources play a dominant role in the observed disparities in college career.

To the best of our knowledge, this is the first attempt to analyse university students' academic trajectories in Italy using national-level data while jointly considering the roles of students' and peers' socioeconomic backgrounds, high school influences and degree programs' characteristics. This contribution is significant given the scarcity of studies that integrate these multifaceted dimensions, despite their acknowledged importance in shaping educational outcomes [1,12]. Overall, our results provide a better understanding of how high schools, degree programs and socioeconomic factors influence university outcomes, offering insights into potential interventions to reduce inequalities in higher education. Although the analysis is limited to students' status one year after enrolment, existing literature consistently shows that first-year outcomes are the strongest predictors of student's likelihood of completing a bachelor's degree [12,22]. The findings align with the international literature, which documents that families' socioeconomic conditions and social ties strongly influence students' educational choices in high school and, indeed, academic orientation and outcomes [27]. A limitation of this study is that the findings on the effects of schools and peers are restricted to the 2019 cohort, which was particularly affected by the CoViD-19 pandemic. However, this cohort is the only one for which we have been able to track students' performance through their second year at university.

CRedit authorship contribution statement

Cristian Usala: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Isabella Sulis:** Writing

– review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Mariano Porcu:** Writing – review & editing, Supervision, Conceptualization.

Data analysis and manipulation

The data used in this study have been processed under the research protocol for the study "From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy" among the Ministry of University and Research, the Ministry of Education and Merit, the University of Palermo as the lead institution, and the INVALSI Institute. The reference researcher is Mariano Porcu.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Isabella Sulis reports financial support was provided by Italian Ministry of Ministry of University and Research, CUP F53D23006150006. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU – Project Title From high school to university: Assessing peers' influence in educational inequalities and performances – CUP F53D23006150006- Grant Assignment Decree No. 1060 adopted on 07/17/2023 by the Italian Ministry of Ministry of University and Research (MUR).

Appendix A

Table A.1

Descriptive statistics on students' characteristics and high schools' self-evaluation exercise by status (2019).

	Regular	Churn	Risk of dropout	Dropout	All
Enrolled	71.9%	4.8%	14.9%	8.4%	115,609
Students' characteristics:					
Average CFU	55.3	19.3	12.6	8.7	43.3
Females (%)	58.2	58.4	47.6	42.7	55.3
Average student ESCS index	0.267	0.410	0.176	0.046	0.241
High school background:					
Average H.S. final mark [60;101]	82.0	79.6	75.7	74.8	80.4
Late diploma (%)	7.0	6.9	14.9	18.6	9.1
HS self eval: Outcomes	0.597	0.609	0.351	0.083	0.518
HS self eval: Processes	0.323	0.350	0.245	0.093	0.293
High school type (by col):					
Liceo Classico (%)	9.8	15.1	9.0	5.1	9.5
Liceo Scientifico (%)	41.2	43.5	37.6	24.7	39.4
Other Institutes (%)	24.3	21.4	19.5	21.2	23.2
Technical Institutes (%)	20.0	15.8	25.9	35.6	22.0
Vocational Institutes (%)	4.8	4.3	8.0	13.4	5.9

Notes: The table shows the descriptive statistics regarding students' characteristics considering the 2019 cohort. It shows the average value for continuous variables (e.g., CFU) and the percentage of students of students in each category for categorical variables (e.g., Females). Columns 1-4 show this information for each student status. Column 5 reports the descriptive statistics considering all students. H.S. indicate the High School. See Table A.2 in the Appendix for the definitions and descriptive statistics of each variable.

Table A.2
Descriptive statistics on variables included in the estimation.

Variables		Descriptive	
Name	Description	Average	Std Dev.
Students' characteristics:			
Female	1 if the student is female	55	
ESCS: Class	INVALSI Economic Social Cultural Status index, class level (only for the 2019 cohort)	0.221	0.458
ESCS: Student	INVALSI Economic Social Cultural Status index, student level (only for the 2019 cohort)	0.241	0.948
Macro area of residence:			
Centre (%)	1 if the student residence municipality is in the Centre	22	
Islands (%)	1 if the student residence municipality is in the Islands	11	
North East (%)	1 if the student residence municipality is in the North East	19	
North West (%)	1 if the student residence municipality is in the North West	23	
South (%)	1 if the student residence municipality is in the South	26	
High school:			
High School final mark	Final high school graduation mark [60;101]	79	12
Not Regular in H.S. (%)	1 if the student was more than 19 years old when graduated from high school	15	
<i>Liceo Classico</i> (%)	1 if the student attended a <i>Liceo Classico</i>	10	
<i>Liceo Scientifico</i> (%)	1 if the student attended a <i>Liceo Scientifico</i>	36	
Technical institute (%)	1 if the student attended a technical institute	23	
Vocational institutes (%)	1 if the student attended a vocational institutes	8	
Other institutes (%)	1 if the student attended a other institutes	23	
Degree programs and university:			
AlmaLaurea: Occupation (%)	Occupation rate one year after graduation from the considered degree program (AlmaLaurea)	36.5	17.8
AlmaLaurea: Wage	Net wage in Euro one year after graduation from the considered degree program (AlmaLaurea)	883.2	243.4
AlmaLaurea: In university (%)	Share of students enrolled at the university one year after graduation from the considered degree program (AlmaLaurea)	41.9	20.5
Revenues per student	Total university revenue from student enrollment divided by the number of enrolled students as measured by USTAT	6,409	3,163
Scholars per student	Scholars per student (USTAT)	0.196	0.049
Non-Scholars per student	Non-Scholars per student (USTAT)	0.182	0.069
Home and university areas:			
Univ. location: Income	Taxable income per capita from the Italian Ministry of Economy and Finance in University city	24,144	3,711
Stud. residence: Unemployment	Unemployment rate in student's province of residence (ISTAT)	0.127	0.063
Stud. residence: Firms	Number of units (firms or branches of firms) in student's municipality of residence	25,068	65,154
Estimated fixed effects:			
Degree program F.E. on CFU	Estimated coefficients of the degree program fixed effects from the first stage	0.000	6.271
High School F.E. on CFU	Estimated coefficients of the high school fixed effects from the first stage	0.000	13
High School self-evaluation:			
HS self eval: Outcomes	High school self evaluation score on outcome estimated using the first component of the Principal Component Analysis (only for the 2019 cohort)	0.518	1.490
HS self eval: Processes	High school self evaluation score on processes estimated using the first component of the Principal Component Analysis (only for the 2019 cohort)	0.293	1.914

The table reports the name and the descriptions of variables used in estimation along with the average and the standard deviation.

Table A.3

Descriptive statistics on the differences between included and excluded records (2019).

	All	Included	Excluded
Enrolled	170,298	67.9%	32.1%
Students' characteristics:			
Average CFU	42.8	41.6	43.3
Females (%)	55.4	55.5	55.3
High school background:			
Average H.S. final mark [60;101]	80.1	79.5	80.4
Late diploma (%)	11.6	16.8	9.1
High school type (by col):			
<i>Liceo Classico (%)</i>	9.6	9.6	9.5
<i>Liceo Scientifico (%)</i>	37.2	32.6	39.4
Other Institutes (%)	24.3	26.8	23.2
Technical Institutes (%)	22.2	22.6	22.0
Vocational Institutes (%)	6.7	8.3	5.9

Notes: Notes: The table shows the descriptive statistics regarding students' characteristics considering the 2019 cohort. It shows the average value for continuous variables (e.g., CFU) and the percentage of students of students in each category for categorical variables (e.g., Females). Column 1 shows the values considering all students. Column 2 reports the information regarding students included in the estimation. Column 3 shows the information for students excluded from the estimation due to missing information. See Table A.2 in the Appendix for the definitions and descriptive statistics of each variable.

Table A.4

Descriptive statistics on the differences between included and excluded records (2019).

ISCED 2-digit	N. Degree programs	Nord-West	Nord-east	Centre	South	Islands
Agriculture	2	2,040	2,473	1,688	2,320	1,438
Architecture and construction	4	1,183	1,565	1,484	1,344	626
Arts	3	2,285	3,305	3,986	2,078	848
Biological and related sciences	2	6,175	5,088	8,114	7,530	2,776
Business and administration	1	24,856	21,127	24,273	24,683	8,373
Education	1	7,226	8,506	10,542	6,019	5,353
Engineering and engineering trades	2	12,931	15,976	15,766	15,410	5,895
Environment	1	3,598	2,756	2,385	1,809	1,003
Health	4	4,149	3,715	4,747	2,974	1,054
Humanities (except languages)	2	3,220	3,845	2,737	1,408	1,320
Information and Communication Technologies (ICTs)	1	7,546	6,526	4,686	7,311	1,966
Inter-disciplinary programmes and qualifications involving agriculture, forestry, fisheries and veterinary	1	2,908	3,655	3,138	2,200	2,261
Journalism and information	1	15,809	6,730	7,397	2,944	3,797
Languages	3	8,078	8,600	8,941	9,543	3,216
Law	1	3,932	3,771	2,457	3,035	654
Mathematics and statistics	2	2,118	2,273	1,850	1,300	490
Personal services	2	4,399	2,849	4,235	3,820	2,646
Physical sciences	6	2,249	1,674	2,596	1,399	1,088
Social and behavioural sciences	7	3,836	4,733	3,647	4,568	2,279
Welfare	1	2,349	2,168	2,809	3,948	1,204

Notes: The table shows the information on the contents covered by the supply of degree programs (*Classi di Laurea*) in Italy. Column 1 shows the ISCED 2-digit classification. Column 2 shows the number of degree programs classified in the ISCED field. Column 3-7 shows the average number of students observed in each degree program in each macro-region.

Data availability

The authors do not have permission to share data.

References

- [1] C. Aina, E. Baici, G. Casalone, F. Pastore, The determinants of university dropout: a review of the socioeconomic literature, *Socioeconomic Planning Sciences* 79 (2022).
- [2] P. Perchinunno, M. Bilancia, D. Vitale, A statistical analysis of factors affecting higher education dropouts, *Social Indicators Research* 156 (2) (2021) 341–362.
- [3] D. Contini, F. Cugnata, A. Scagni, Social selection in higher education. Enrolment, dropout and timely degree attainment in Italy, *Higher Education* 75 (2018) 785–808.
- [4] F. Belloc, A. Maruotti, L. Petrella, How individual characteristics affect university students dropout: a semiparametric mixed-effects model for an Italian case study, *Journal of Applied Statistics* 38 (10) (2011) 2225–2239.
- [5] S. Meggiolaro, A. Giraldo, R. Clerici, A multilevel competing risks model for analysis of university students' careers in Italy, *Studies in Higher Education* 42 (2017) 1259–1274.
- [6] OECD, Education at a Glance 2022: OECD Indicators, OECD Publishing, Paris, 2022.
- [7] D. Contini, R. Zotti, Do financial conditions play a role in university dropout? New evidence from administrative data, in: D. Checchi, T. Jappelli, A. Uricchio (Eds.), *Teaching, Research and Academic Careers*, Springer, Cham, 2022.
- [8] Cristian Usala, Isabella Sulis, Luisa Salaris, Nicola Tedesco, Mariano Porcu, Barbara Barbieri, Disentangling the secondary school effects in the assessment of university inefficiencies, *Studies in Higher Education* (2024) 1–21, <https://doi.org/10.1080/03075079.2024.2345188>.
- [9] United Nations, Transforming our world: the 2030 Agenda for Sustainable Development, <https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>, 2015.
- [10] A. Busetta, E. Fabrizi, I. Sulis, G. Ragozini, et al., (Im) mobilità sociale delle famiglie, in: *Rapporto sulla popolazione Le famiglie in Italia. Forme, ostacoli, sfide*, Società Editrice il Mulino spa, 2023.
- [11] Umut Türk, Socio-economic determinants of student mobility and inequality of access to higher education in Italy, *Networks and Spatial Economics* (ISSN 1572-9427) 19 (1) (2019) 125–148, <https://doi.org/10.1007/s11067-019-09445-w>.
- [12] D. Contini, G. Salza, Too few university graduates: inclusiveness and effectiveness of the Italian higher education system, *Socio-Economic Planning Sciences* 71 (2020) 100803, <https://doi.org/10.1016/j.seps.2020.100803>.
- [13] Gary S. Becker, Investment in human capital: a theoretical analysis, *Journal of Political Economy* 70 (5, Part 2) (1962) 9–49.
- [14] Beverly Duncan, Dropouts and the unemployed, *Journal of Political Economy* 73 (2) (1965) 121–134.
- [15] Daniel I. Rees, H. Naci Mocan, Labor market conditions and the high school dropout rate: evidence from New York State, *Economics of Education Review* 16 (2) (1997) 103–109.
- [16] G. Di Pietro, Regional labour market conditions and university dropout rates: evidence from Italy, *Regional Studies* 40 (6) (2006) 617–630.
- [17] Vincent Tinto, Dropout from higher education: a theoretical synthesis of recent research, *Review of Educational Research* 45 (1) (1975) 89–125, <https://doi.org/10.3102/00346543045001089>.
- [18] Ernest T. Pascarella, Patrick T. Terenzini, Student-faculty informal relationships and freshman year educational outcomes, *The Journal of Educational Research* 71 (4) (1978) 183–189.
- [19] Ralph Stinebrickner, Todd R. Stinebrickner, What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds, *Journal of Public Economics* 90 (8–9) (2006) 1435–1454.
- [20] Bruce Sacerdote, Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? Vol. 3. 11, 1st ed., Elsevier B.V., 2011, pp. 249–277.
- [21] MOBISU.IT Database [Mobilità degli Studi Universitari in Italia], Research protocol MUR - Universities of Cagliari, Palermo, Siena, Torino, Sassari, Firenze, Cattolica and Napoli Federico II, Scientific Coordinator Massimo Attanasio (UNIPA), Data Source ANS-MUR/CINECA, 2022.
- [22] A. Priulla, N. D'Angelo, M. Attanasio, An analysis of Italian university students' performance through segmented regression models: gender differences in STEM courses, *Genus* 77 (2021) 11.
- [23] Cristian Usala, Ilaria Primerano, Francesco Santelli, Giancarlo Ragozini, The more the better? How degree programs' variety affects university students' churn risk, *Socio-Economic Planning Sciences* (ISSN 0038-0121) 94 (2024) 101926, <https://doi.org/10.1016/j.seps.2024.101926>.
- [24] Julia Horstschräer, University rankings in action? The importance of rankings and an excellence competition for university choice of high-ability students, *Economics of Education Review* (ISSN 0272-7757) 31 (6) (2012) 1162–1176, <https://doi.org/10.1016/j.econedurev.2012.07.018>.
- [25] Mark E. Engberg, Gregory C. Wolniak, Examining the effects of high school contexts on postsecondary enrollment, *Research in Higher Education* 51 (2010) 132–153.
- [26] Mark E. Engberg, Gregory C. Wolniak, An examination of the moderating effects of the high school socioeconomic context on college enrollment, *The High School Journal* 97 (4) (2014) 240–263.
- [27] Gregory C. Wolniak, Mark E. Engberg, Academic achievement in the first year of college: evidence of the pervasive effects of the high school context, *Research in Higher Education* 51 (5) (2010) 451–467, <https://doi.org/10.1007/s11162-010-9165-4>.