n. 59 (04/2019)

The contribution of high schools to university students' academic performance: The case of Eduscopio

di Carmen Aina, Massimiliano Bratti, Enrico Lippo

Fondazione Agnelli

The contribution of high schools to university students' academic performance: The case of Eduscopio*

Carmen Aina† Massimiliano Bratti‡ Enrico Lippo§

Abstract

The Eduscopio project is the first attempt to evaluate the contribution of Italian high schools to university students' academic performance, based on their first-year of enrollment. After describing the main features of the school performance indicators produced by Eduscopio, we check their robustness by computing indicators based on the performance achieved not only in the first-year but also on cumulative performance at the third-year of university enrollment, and on longer term outcomes, such as the probability of graduation within the degree legal duration. Our analysis demonstrates that the contribution of the secondary school of origin is not temporary and limited to the first-year but also plays a significant role in the subsequent academic years, and that the high schools' ranking based on the different indicators remains rather stable. Our estimates confirm a high positive correlation between the first- and the third-year student achievement, as well as between the first-year student performance and the probability of on-time graduation.

Keywords: Eduscopio, high school, university performance, school ranking, Italy
JEL codes: I20; I21; I23

*Funding from the Fondazione Giovanni Agnelli (FGA) is gratefully acknowledged. The authors are grateful to the FGA for making available the data of the National Archive of Students and Graduates (*Archivio Nazionale degli Studenti e dei Laureati*, ANS) for research purposes. Comments from Martino Bernardi, Gianfranco De Simone, Andrea Gavosto and Angela Martini are gratefully acknowledged. All authors disclaim that they have no conflict of interest. The usual disclaimers apply.

†(corresponding author) Università del Piemonte Orientale (Novara, Italy); GLO; carmen.aina@uniupo.it

‡Università degli Studi di Milano (Milan, Italy); IZA (Bonn, Germany); Centro Studi Luca d'Agliano (Milan, Italy); GLO; massimiliano.bratti@unimi.it

§Università del Piemonte Orientale (Novara, Italy); enrico.lippo@uniupo.it

1 Introduction

Making conscious choices requires having information on the costs and benefits of each available alternative. However, such information is often unavailable. Until very recently, this was the case for the choice of high school in Italy: there was no rigorous assessment of school performance, so parents and students had to rely mainly on informal knowledge provided by other parents or older peers. Although from the academic year 2009/2010 a standardized test for 10th graders, i.e. the second year of high school, was introduced in Italy — the so-called INVALSI (i.e. National Institute for the Evaluation of the Education System) test — average results were not publicly released for all schools.[1] Information was sparse because the decision to make the results public was left to the individual initiative of school principals.[2]

Partly to address this lack of information, the *Fondazione Giovanni Agnelli* (hereafter FGA) launched in November 2014 a new web portal named Eduscopio,[3] listing school performance indicators (PIs) for Italian high schools. The main aim of Eduscopio is to build PIs to measure the contribution of each school to university performance using the *first-year performance* of all students coming from that school (*Eduscopio Università*), or to students' labor market outcomes after graduating from high school (*Eduscopio Lavoro*). The first set of PIs is mainly thought for academic and technical schools, and the second one for technical and vocational schools. In the current paper, we focus on the *Eduscopio Università*'s (Eduscopio, for brevity) PIs , which are computed using administrative data on the Italian population of students enrolled in Higher Education (HE, hereafter) in Italy, which are gathered by the National Archive of Students and Graduates (henceforth ANS) maintained by the Italian Ministry of Education, University and Research (MIUR). Although Eduscopio

[1] Here we refer to "high school" and upper secondary education (spanning the 9th-13th grades) as synonyms. Italy has a tracked upper secondary school system, and the three main tracks are: academic high schools (*licei*), technical schools (*istituti tecnici*), and vocational schools (*istituti professionali*), which are further divided by field of study (see Section 2.2).

[2] Moreover, school principals generally published the average raw score, without normalizing it for the student characteristics.

[3] See for details: `https://eduscopio.it/`.

reports information on a very high number of high schools across the whole Italy, the web portal advises users to make comparisons only within the same school track and between geographically close schools (e.g. located in the same city or province).

Eduscopio received good press coverage from its inception. Many national newspapers[4] have devoted space to disseminating the main results of Eduscopio. This in turn has generated great attention and discussion on the local news, putting pressure on the principals of the schools with the poorest performance. Since the first release (November 2014), Eduscopio has seen a sharp rise in the interest of users. In December 2015, one year after the first Eduscopio's release, the hits on the website increased by about 13.6%, especially those made by distinct users (18.3%), as reported in Vuri (2018).

In this study, we assess the robustness of the methodology employed by FGA to build Eduscopio's indicators. This is important given the growing relevance of Eduscopio as a source of information for school choice in the Italian context (Vuri, 2018). First, we extend the time horizon of the ANS data used to compute the indicators, considering students' performance both one year and three years from first enrollment. Indeed, FGA originally chose to compute the indicators on the first year only, on the basis of the strong correlation of the latter with later student performance (see, for instance, Porter and Swing, 2006; Lang, 2007; Jamelske, 2009). However, students with certain school backgrounds may systematically face problems at the start of their academic careers but catch up later. Such schools would be penalized by how the current Eduscopio's indicators are built. Second, we compute school PIs based on the probability of graduating within the legal duration of the degree course. The latter is a meaningful PI given the long average graduation times observed in Italy (Aina et al., 2011; Garibaldi et al., 2012).

Our analysis demonstrates that Eduscopio's PIs are informative on the "quality"[5] of high schools in the Italian context. In particular, the high school ranking built on first-year

[4]For example *Il Corriere della Sera*, *La Repubblica*, *Il Sole 24 ore*, *La Stampa*, *Libero*, among others.

[5] In what follows, we sometimes refer to "school quality" in a loose sense meaning everything at the school level that can contribute to a student's future academic and potentially labor market success.

performance is robust to measuring performance at the third-year after first enrollment and to considering on-time graduation.

The structure of the paper is as follows. Section 2 describes the main characteristics of the data, and the sample selections made before computing the indicators, which mimic the selections made by FGA to build Eduscopio. Section 3 presents the methodology used by Eduscopio, and how the proposed indicators are extended to reflect longer-term student outcomes, namely academic performance within three years from first-time university enrollment, and the probability of graduation within the degree legal duration. Our main results are discussed in section 4 while some robustness checks are reported in the Appendix B. Section 5 draws conclusions.

2 Data and sample selection

2.1 Eduscopio data

The dataset used in our empirical exercise is built by linking three distinct administrative sources of data provided by MIUR, namely:

a. the National Archive of Schools (i.e. *Anagrafe Nazionale delle Scuole*);

b. the dataset containing details on the characteristics of the high schools (i.e. *Scuola in Chiaro*);

c. the National Archive of Students and Graduates (i.e. *Archivio Nazionale degli Studenti e dei Laureati*, ANS).

From the National Archive of Schools, we drew the list of the Italian high schools including the name, the type (i.e. school track) and the address. To this set of variables, we added the number of secondary school graduates and their average final graduation mark by type of high school. By exploiting such information, we are able to select the academic and technical high schools only, which are the focus of the *Eduscopio Università* project. The academic

careers of the students enrolled in the Italian tertiary education system, instead, are drawn from the ANS. ANS contains the administrative records of each student, so that, during the academic years 2009/10, 2010/11 and 2011/12, for each cohort of freshmen we have details on gender, nationality, year of birth, high school's municipality, name and type of high school, year of high school diploma, high school final mark, the list of university exams passed and the corresponding amount of credits, the dates when exams were passed, grade per exam, name of the degree course (along with college major) and of the University, and part-time student status (i.e. part-time or full-time). The academic career of each student is available for the entire period she is included in the ANS archive, also in case she changed degree course and/or University, up to the 31st of October 2017. The sample does not cover individuals enrolled in tertiary education courses provided by Higher Colleges of Technology, by Advanced Training Schools in Art and Music subjects, and by foreign HE Institutions.

2.2 Sample selection criteria

The ANS data allow to track the careers of the individuals who enrolled in the Italian HE system, either in a Bachelor's or in a single-cycle degree (i.e. *Laurea a Ciclo Unico*),[6] in the academic years 2009/2010, 2010/2011 and 2011/2012. Each student's career is then observed from the matriculation day up to its end (i.e. graduation, dropout) or, in case a student is still enrolled in HE, up to 31st of October 2017. Overall, we observe 1,094,875 careers but, considering that a student may have more than one career, our sample consists of 1,026,111 distinct students and represents the population of individuals who enrolled in the university system in the aforementioned academic years. Starting from this dataset, to increase comparability, we keep only those students who received their diploma from an Italian high school in the school years relevant for our analysis, which are 2008/2009,

[6] After the Bologna process, in Italy HE consists of two levels of degree: first-level degrees, whose duration is generally three years and second-level degrees (2 years). However, in some degree subjects university degrees are structured in a single cycle, lasting 5 years or more (i.e. Medicine and Surgery, Dentistry and Dental Prosthesis, Pharmacy, Law, Primary Education Sciences, Veterinary, Architecture, Pharmaceutical Chemistry and Technology).

2009/2010 and 2010/2011. Therefore, in our sample, an individual may have started her academic career with none, one or two years of delay. Moreover, we drop those students that were more than 22 years old when they achieved secondary school diploma. This latter choice is due to the fact that students with delayed careers (e.g., those who retained grades) typically show fragmented careers in different schools. As a consequence, it would be incorrect to entirely attribute the school effect to the institution in which they are awarded their diploma. Finally, we exclude from our analysis students who graduated in Professional Institutes (4.29%), Applied Art High Schools (0.66%) or Art Lyceum (1.36%), which are a mix of vocational (the first two) and academic (the last one) schools.

After these selections, our working sample consists of 692,746 careers belonging to 643,867 students. More in detail, 596,625 students (92.66%) show a single university career, 45,638 students (7.09%) have two careers, and the remaining part shows from three up to five careers (0.24% of the sample corresponding to 1,604 individuals). In order to simplify our analysis, we drop the students who have more than two university careers, remaining with a sample of 642,263 individuals. Each student of our working sample is then linked with her high school of origin using the National Archive of Schools and *Scuola in Chiaro*.

The Italian secondary school system, as designed by the 2010 reform, is organized into three main types of high schools, which are subsequently divided into further specializations: (A) Lyceum *(Liceo)*; (B) Technical Institutes *(Istituti Tecnici)* and (C) Professional Institutes *(Istituti Professionali)*. Lyceum is further divided into six types: (1) Art Lyceum *(Liceo Artistico)*; (2) Classic Lyceum *(Liceo Classico)*; (3) Linguistic Lyceum *(Liceo Linguistico)*; (4) Music and Dance Lyceum *(Liceo Musicale e Coreutico)*; (5) Scientific Lyceum *(Liceo Scientifico)* and (6) Social Sciences Lyceum *(Liceo delle Scienze Umane)*. Technical Institutes are in turn organized into two main macro-fields and eleven sub-fields: (1) Economic Sector *(Settore Economico)* which has 2 sub-fields and (2) Technological Sector *(Settore Tecnologico)* with 9 sub-fields. Finally, Professional Institutes *(Istituti Professionali)* are organized into two macro-areas and 6 sub-fields: (1) Services Sector *(Settore dei Servizi)*

with 4 sub-fields and (2) Industry and Handicraft Sector *(Settore Industria e Artigianato)* with 2 sub-fields.

We are not able to include in our analysis the Art Lyceum and the Music and Dance Lyceum due to the fact that most of their students enroll in the AFAM system (Advanced Training Schools in Art and Music subjects) which is not covered by the ANS data.

Moreover, we exclude from the analysis Professional Institutes, which are typically oriented towards vocational subjects and, having a job-oriented education, also have low university enrollment rates. Consequently, we focus on the remaining six types of high schools: (1) Classic Lyceum; (2) Linguistic Lyceum; (3) Scientific Lyceum; (4) Social Sciences Lyceum; (5) Technical-Economic Sector Institutes and (6) Technical-Technological Sector Institutes. We prefer to keep the distinction between Technological and Economic schools for the Technical Institutes because, although they are both classified as Technical Institutes by the reform, these schools are historically considered as belonging to very different fields of study. Among these remaining types of high schools, we exclude from the analysis those schools having an aggregate university matriculation rate lower than 33% and with less than 21 students (the size of around a school class) enrolled in HE considering the three cohorts of freshmen included in the analysis. Although these thresholds are somehow arbitrary, they are in line with the sample selection made in the original version of Eduscopio, and have the objective of relying on a sufficient number of students for each school to build the Eduscopio's PIs.

The final sample is composed of 642,263 individuals, whereas the schools ranked by Eduscopio are 4,280. The descriptive statistics of student characteristics are reported in Table 1. The majority of students are females (about 57%) and Italian citizens (98%). The average age at university matriculation is 19 and the mean of the high school mark is 79.[7] Students tend to enroll at university the same year in which they complete upper secondary education. Only 8.46% and 0.4% take one- and two-year gaps, respectively. The

[7] The high school makr ranges between 60 to 100 *cum laude*. We assigned the value 101 to 100 *cum laude*.

percentages of students by region reflect the size of regional populations. Students are not very mobile as they attend degree courses, on average, 86 kilometers away from the high school they attended. Table 2 reports the characteristics of the high schools ranked by Eduscopio. About 70% of them are academic (i.e. *Lyceum*), whereof 30.49% are Scientific, and 13% are private institutions. The average 13th graders' cohort size is of 196 students, while the mean enrollment rate in HE is 71.68%. The majority of high school graduates are females (57.53%), whereas only few students are foreigners (2.54%).

2.3 Outcome variables

Using information on individuals' university careers, Eduscopio builds, for each student, two academic performance indicators:

a. the Percentage of University Credits (PUC);

b. the Grade Point Average (GPA).

Both indicators are computed at the end of the first academic year and used as dependent variables in the regressions outlined in Section 3.

The PUC achieved by a student at the end of the first academic year, which is conventionally set at the 30th April $t+2$, where t is the calendar year of the student's first-time enrollment, is defined as:

$$PUC_t^{ij} = \frac{\sum_{k=1}^{n_i} CFU_k^{ij}}{\overline{CFU}^{ij}} \tag{1}$$

where i indexes students, j high schools, and t the calendar year of enrollment; n_i is the number of exams passed by individual i, k is the exam subscript, and $\overline{CFU}^{ij}$ is the number of achievable university credits which depends both on the major and on student status (i.e. part-time or full-time).[8]

[8] In the Italian HE system each exam is associated with a given number of CFUs, and each CFU corresponds to 25 hours of student workload (including individual study hours). A first-level degree requires 180 CFUs.

Student's GPA at the end of the first academic year is defined as:

$$GPA_t^{ij} = \frac{\sum_{k=1}^{n_i} mark_k^{ij} \cdot CFU_k^{ij}}{\sum_{k=1}^{n_i} CFU_k^{ij}} \tag{2}$$

where each mark obtained by student i for exam k is weighted by the corresponding number of credits for that exam (CFU_k^{ij}). Notice that $\sum_{k=1}^{n_i} CFU_k^{ij}$ may be different from $\overline{CFU}^{ij}$ because some exams are awarded a "pass" grade.[9]

In order to check the robustness of the ranking obtained by Eduscopio, we compute the PUC and GPA indicators also at the end of the third academic year, which we conventionally set at the 30th of April of year $t+4$. We also investigate the robustness of the indicators considering the 31th October of year $t+1$ and $t+3$ as the cut-off dates. The latter results are reported in the Appendix B.

As a further PI, we look at a student's graduation status, namely we create a dummy variable which equals one if a student obtains a degree within the degree legal duration and zero otherwise:[10]

$$D_t^{ij} = I(\text{graduation on time}) \tag{3}$$

where $I(.)$ is the indicator function. $D_t^{ij} = 0$ includes both individuals who dropped out from tertiary education and those who did not graduate on time or still have to graduate.

3 Eduscopio: The methodology

The procedure followed by FGA to compute school rankings consists of three steps (Bernardi and De Simone, 2018):

Step 1. The outcome variables defined in Section 2.3 are regressed on some student and

[9] If a student does not pass any exam in the selected time period her indicators will be PUC=0 (i.e. inactive students) with missing GPA. Following Eduscopio's methodology, inactive students only contribute to the PUC part of the index (see Step 3 in the next section).

[10] The legal duration is equal to three years for first-level degrees and five or six years (Medical Studies) for single-cycle degrees.

university level characteristics to "clean" them from the average differences in student performance that are university- and college-major specific;

Step 2. The raw outcome variables are "standardized" by subtracting the university-college-major fixed effects that are estimated in the previous step and then they are regressed on a set of school fixed effects and year of secondary school graduation fixed effects. This procedure is applied to both the percentage of university credits (PUC) and the grade point average (GPA), which are the two building blocks of the Eduscopio index;

Step 3. The school fixed effects for PUC and GPA estimated in the previous step are normalized to vary in the 0-100 scale and then combined together to build the Eduscopio final index.

In the following sections, we describe the three steps in more detail.

3.1 Step 1. Regression of student performance on individual-level and university-level characteristics

The first step can be described by the following equation

$$y_{ict}^{p} = \beta_0 + \alpha_0 + \beta_1 \mathbf{X}_i + \beta_2 \mathbf{Z}_i + \sum\nolimits_c \phi_c + \sum\nolimits_t \tau_t + \epsilon_{ict} \tag{4}$$

where i, c and t are individual, higher education institution-college major and time subscripts, respectively; p = (PUC, GPA) is the outcome superscript. It is worth noting that we use pooled cross-section data, i.e. we only have one observation for each individual. However, there is a time dimension since we pool three cohorts of university entrants in the analysis. y_{ict}^{p} is the outcome of interest (PUC, GPA). $\mathbf{X}_i$ is a vector of individual level characteristics including: gender, immigrant status, distance between the municipality of the secondary school and that of the university degree course attended.[11] $\mathbf{Z}_i$ is a vector of characteristics related to the student's secondary school career: school track, high school final mark, region,

[11] These geodesic distances are computed using municipalities' centroids (for both the high school and the university of enrollment) as input coordinates and the Stata command *geodist*.

age at high school completion and age at university matriculation. ϕ_c are university-college major fixed effects, which control for average differences in grading policies or the speed of student progression across higher education institutions and college majors (see, for instance, Bagues et al., 2008). τ_t are academic year of enrollment fixed effects. ϵ_{ict} is an individual error term.

The main purpose of this step is to estimate a set of university-college major fixed effects (ϕ_c) that reflects differences in grading styles or standards that are university-college major specific. To this end, it is crucial to control for students' characteristics that may affect university performance such as, for instance, secondary school track or high school final mark, which are proxies of a student's ability.

The estimates from equation (4) are used to compute the following net (or standardized) performance, both for PUC and GPA

$$\tilde{y}_{it}^{k} = y_{ict} - \sum\nolimits_{c} \phi_c. \tag{5}$$

In this way, we obtain for each student a measure of academic performance that is standardized with respect to the "average academic performance" of her peers enrolled in the same college major and *Alma Mater*.

3.2 Step 2. Estimation of school fixed effects

The standardized PIs ($\tilde{y}_{it}^{k}$s) are regressed on a set of school (θ_j) and year of matriculation (π_t) fixed effects, as follows

$$\tilde{y}_{ijt}^{p} = \mu_0 + \sum \theta_j + \sum \pi_t + \rho_{ijt} \tag{6}$$

where i, j and t are individual, high school and time subscripts, respectively; p = (PUC, GPA) is the outcome superscript, and ρ_{ijt} is an error term. Our main coefficients of interest are the school fixed effects, which capture the average performance of all students enrolled in HE coming from the same school. Such average performance also reflects the contribution of

the school track and of other measures of school quality which are captured by the average secondary school final mark awarded by that school.[12]

It is worth noting that these school fixed effects capture a mix of different effects that schools have on students' university performance. In particular, they capture both the impact of average students' individual ability and of the average socio-economic status of the students enrolled in each school as well as the school value added (i.e. the school contribution to academic performance, over and above student characteristics). Indeed, FGA claims that parents are interested in all these aspects when choosing a high school for their children.[13]

3.3 Step 3. Construction of the Eduscopio index

The fixed effects computed in Section 3.2 are first re-scaled between 0 and 100

$$\theta^p_{[0,100],j} = \frac{\theta^p_j - \theta^p_{min}}{\theta^p_{min} - \theta^p_{max}} \cdot 100 \tag{7}$$

where θ^p_{min} and θ^p_{max} are the smallest and the largest fixed effects estimated for outcome p, respectively. Then the two rescaled fixed effects for CFUs and average exam grades are

[12] The rationale for not purging the residual of equation (5) from the correlations with the school track and average secondary school marks is that they partly reflect the school contribution to the students' academic readiness before enrolling at university.

[13] Since the goal of this paper is to check the robustness of the Eduscopio's methodology, we abstract here from in-depth considerations about the methodology. In brief, the way the indicators were built in the past suffers from the lack of standardized test scores in the Italian context. The educational production function could be specified as $y_i = f(a_i, a_{-i}, S_i, X_i)$, where a_i is individual i ability, a_{-i} is her peers' ability, S_i are net school effects (i.e. net of individual and peer ability, i.e. school value added), X_i are other individual characteristics (such as socio-economic status, SES). In the absence of information on a_i and a_{-i} at high school entry, or even of proxies of it (e.g. SES) the estimated (gross) school effects also capture the effect of students' own ability. FGA is aware of this, and indeed allows to make comparisons across schools belonging to the same school track and that are near geographically. The idea is that applying these criteria should limit the heterogeneity in (unobserved) average student ability across schools, and the comparison is then informative of the net school effects. In case this assumption fails, Eduscopio would remain informative for both HE Institutions, who may be interested in selecting the students with best potential for HE, and employers, who may not care about the source of workers' higher productivity, but less for families. The latter, in order to choose the best school for their children must be able to factor out the part of the school effect that is accounted for by the average ability of the school intake (i.e. to know if their children, conditional on their ability levels, if enrolled in a given school will have similarly good results at university as the students who enrolled in the past in the same school). Matching the National Archive of Schools, ANS and *Scuola in Chiaro* with the Invalsi standardized test scores before high school entry, which are available for the most recent student cohorts, could overcome many of these methodological concerns.

combined into the Eduscopio index

$$\text{ES}_j = 0.5 \cdot \theta^{\text{cfu}}_{[0,100],j} + 0.5 \cdot \theta^{\text{grades}}_{[0,100],j}. \tag{8}$$

which weights equally PUC and GPA. This is the way the Eduscopio index is computed for first-year performance, and how we built it for three-year performance. For graduation outcomes, instead, we built the Eduscopio index as the normalized school fixed effects (in the 0-100 range) computed from a linear probability model for on-time graduation.

For the construction of the confidence intervals of the Eduscopio index we followed the procedure described in the Eduscopio's technical report (Bernardi and De Simone, 2018), which is also reported in Appendix A.

4 Main results

4.1 Step 1 results

Results of the equation (4) for the outcomes described in section 2.3 are reported in Tables 3, 4 and 5. For each outcome, we provide three specifications. In particular, in the baseline specification (i.e. Model 1) we control only for individual characteristics, high school characteristics and academic year of enrollment; whereas in Model 2 we add the university-college major fixed effects, and finally in Model 3 the interaction between high-school and its region is included.

Concerning the first-year performance GPA and PUC (see Table 3), we notice that results for the high school final mark, academic year of matriculation and studying in a private secondary school are stable across the three models. To be more precise, for each additional year of age at high school graduation the PUC decreases by about 5 percentage points,[14] while each additional point in the high school mark raises the percentage of first-year credits by about 1 percentage point. By contrast, students who obtained the high school diploma

[14] Since the PUC varies between 0 and 1, the regression coefficients must be multiplied by 100 to obtain percentage point changes.

from a private school face a penalty in the first-year credits performance of about 11-12 percentage points.[15] Foreign students earn less credits than their Italian peers, but this gap becomes smaller when the type of degree course and tertiary institution is controlled for. Females, on average, acquire more credits than males, but after controlling for the *Alma Mater* and college major this advantage halves (i.e. from 3.7 to 1.6 percentage points). This latter finding reflects gender differences in the choice of field of study, with women less prone to enroll in Science, Technology, Engineering and Mathematics (STEM) degrees, which are generally more academically demanding than degrees in other majors (see Griffith, 2010; Riegle-Crumb et al., 2012; Card and Payne, 2017). The PUC obtained during the first university year improves as the distance between the high school and the university increases, in particular for each extra 100 kilometers the percentage of credits increases of 1.8 percentage points (model 1), but again in the full specification this advantage becomes smaller (0.6 percentage points). This result may capture the higher effort and motivation of individuals who migrate for study purposes. Results for first-year GPA are qualitatively similar. GPA increases when the freshman is female, has a higher diploma final mark, graduated in a public secondary school, studied closer to the high school's municipality and is an Italian national. Again, controlling for college major and HE institution of enrollment (Model 3), we notice that women's advantage in GPA decreases and is almost 0.08 points, while in the baseline specification, in which these variables are omitted, this advantage is much larger, of about 0.5 points, confirming that women tend to enroll in less demanding fields. High school final mark is a significant predictor of GPA with a 10-point increase in the average grade received at the end of secondary school being associated with about a one point increase in the GPA. Unlike with PUC, studying in a university farther away from the student's high school of origin reduces the GPA. These results suggest that there is a potential trade-off between the quantity of exams that students pass in the first-year and the quality of their careers (measured in terms of GPA), and that students enrolled in a

[15] This is not surprising in the Italian context in which many private schools are "remedial schools" (Bertola et al., 2007).

degree program farther from their geographical area of origin could give more importance to on-time completion of the degree course than to average grades. This is consistent with the higher educational costs they face. Interestingly, the year of enrollment dummies show a progressive improvement of PUC and deterioration of GPA across cohorts.

To analyze whether the role played by all the variables included in equation (4) is persistent over the student's academic career, we estimate models with the PUC and GPA measured three years after enrollment. These estimates are reported in Table 4. Results show the same pattern as for the first-year performance, and the estimates are quantitatively similar and show similar levels of statistical significance.

The correlations in models 1 and 2 are only partially attenuated in magnitude when university-college major fixed effects are included to clean the estimates from possible differences in grading standards across universities and college majors.

Finally, we run the three specifications on the last outcome, namely on-time graduation.[16] Table 5 reports comparable correlations between the covariates and the outcome as shown for first-year PUC. Chiefly, a one point increase in the high school final mark provides a 1 percentage point advantage in on-time graduation, while attending a private high school entails a reduction in this probability of about 10 percentage points (model 3). The estimates confirm the positive correlation between distance and the speed of students' academic careers. Indeed, for every extra 100 kilometers of distance the likelihood of getting the degree within the legal duration increases of about 1.3 percentage points (model 3). The effect of distance is inverse-U shaped and reaches a maximum at 475 kilometers of distance from the high school attended.

[16] Note that for this set of estimates the sample size is smaller as we do not include students enrolled in Medical studies during the academic year 2010/2011 and 2011/2012 as for them the time span of our administrative data is not long enough to allow them to graduate within the legal duration.

4.2 Step 2 results

Figures 1 and 2 report the estimates of the school fixed effects obtained from equation (6) for the two first-year outcomes, namely PUC and GPA. High school FEs are ordered from the highest to the lowest value. Considering both outcomes, it emerges that for the majority of high schools (in the center of the distribution) the academic performance is statistically indistinguishable (considering the overlapping confidence intervals). This pattern partly reflects the public service nature of the Italian secondary school system, which should supply quite a homogenous good in terms of quality.

Similarly, once the third-year academic outcomes are plotted (see Figures 3 and 4), the high schools' contributions to each outcome are comparable with the pictures shown for the first academic year.

Results are also confirmed when the high school ranking is based on the probability of on-time graduation (Figure 5).

4.3 Step 3 results

The high school ranking for the first- and third-year Eduscopio indexes are plotted in Figures 6 and 7, respectively, while for on-time graduation the ranking is reported in Figure 8.

To check the robustness of the ranking to changing the time at which academic performance is measured, we regressed the first-year combined performance of PUC and GPA, that is the first-year Eduscopio index, on the corresponding third-year index and on the graduation on time index, respectively. Figure 9 shows that the correlation between the first- and third-year Eduscopio index is very strong ($\beta = 1.05$), which suggests that the benefits associated to the high school of origin do not disappear over time, but still continue to play a role on the years subsequent to matriculation. Similarly, the correlation between the first-year Eduscopio index and graduation within the legal length index confirms the high school's persistent effect (Figure 10), as the first-year Eduscopio index is strongly associated with the one relating to the probability of completing the academic studies on time ($\beta = 0.94$).

These patterns are also confirmed when we consider these indicators by school track (Figures 11 and 12). The correlation is smaller for Technical Institutes in the Technological track though, especially when we consider the association between the first-year Eduscopio index and on-time graduation ($\beta = 0.69$).

Finally, in Table 7 we report the estimates of the probability of changing the position in the school ranking deciles (i.e. transition matrix) by the high school's characteristics, such as school track and status (public vs. private) when the time-span in which student performance is measured increases (from one to three years after first university enrollment). With reference to the first column, we notice that the probability of remaining in the main diagonal of the transition matrix is lower for students that do not come from academic tracks, especially students of *TI-Technological* schools (13 percentage points), followed immediately by students with a secondary diploma awarded by a *Linguistic Lyceum* or a *TI-Economics*. Indeed, students graduated from a *Scientific lyceum* track are not statistically different from those coming from *Classic Lyceum*. In the second column of Table 7, we analyze a school's probability of improving its performance between the ranking based on first-year performance and the one based on third-year performance. Results suggest that only students from the *Social Sciences Lyceum* have a higher probability of improving their ranking, of about 5.6 percentage points, while high school graduates not coming from academic tracks and *Scientific Lyceum* are less likely to improve their position. About the probability of performing worse in the ranking when increasing the time-span of the academic career considered in the PIs (column 3), we notice that the likelihood of moving below the diagonal, compared to students from *Classic Lyceum*, is higher for any type of high school, especially for students from a non-academic track.

Since comparisons in Eduscopio should be made across schools within the same track and that are geographically close, Table 7 may not be very informative. For this reason in the Appendix C we have reported some additional tables and figures for three large Italian cities: Rome, Milan and Naples.

Table C1 reports the first 10 and last 10 academic track schools for the city of Rome listed in decreasing order of the first-year Eduscopio index. The table shows the rank, the value of the index and the corresponding confidence interval (CI) in the columns (1), (2) and (3), respectively. Columns (4), (5) and (6) report the same variables but for the third-year Eduscopio index. A first thing worth noting is that although the CIs of the first (last) 10 schools usually overlap, and therefore many of these schools cannot be ranked in statistical terms, the same is not true when making comparisons between the first 10 and the last 10 schools. Hence, even considering a school track that is supposed to prepare most students for HE, it is possible to observe sharp differences in the first-year university performance of the students coming from different schools.[17] Comparing the change in the rank between the first- and the third-year Eduscopio index, we see limited shifts in the ordering of schools. Tables C2, reporting the results for the Human Sciences and Linguistic tracks, and C3, reporting the results for the technical track convey a similar picture. Perhaps, what may appear surprising is that the range of the Eduscopio index seems to be wider for the academic track than for the other two track groups, as shown by the gap in the index values between the schools at the top and at the bottom of the ranking. This may be partly explained by the higher number of academic schools satisfying the selection criteria listed in Section 2.2 and entering our estimation sample, which may generate a higher heterogeneity across schools in Rome.

These rankings are visualized in Figure C1, which reports the first-year Eduscopio index point estimates with CIs by school track for the city of Rome. The same is done for the cities of Milan and Naples, in figures C2 and C3, respectively. These figures show, for instance, the higher dispersion in the value of the index for the academic track schools in Rome compared to Milan.

[17] These differences are not related to the potential differences in the college majors chosen by students coming from the different schools (and the related grading standards/practices) because they have been controlled for in the second step of the estimation (Section 3.1).

5 Concluding remarks

Eduscopio is a web-portal recently created by the Fondazione Giovanni Agnelli (FGA) providing high school performance indicators for almost all Italian academic and technical schools. High schools are ranked according to the university performance of the students who enrolled in Italian HE Institutions. In particular, the Eduscopio index takes into account the percentage of university credits (PUC) and the grade point average (GPA) of students in their first-year after enrollment.

Given the increasing importance of Eduscopio as a potential source of information to support Italian students and their families (Vuri, 2018) in the choice of the high school, in this paper we carry out some robustness checks on the methodology used to build the Eduscopio index. In particular, we extend the time-span in which student performance is measured, from the first- to the third-year after student enrollment, and we also build an index based on the probability of on-time graduation. The latter is a meaningful performance indicator given the high percentage of university students who drop out or graduate with significant delays in Italy. Both robustness checks address the potential concern that students coming from certain schools may have a "late start", in terms of academic performance, but catch up later. In fact, we show that the correlation between indexes built on first-year and third-year performance is very high, both in general and by school type. A similarly high correlation is also observed between the first-year index and the on-time graduation index.

The conclusion of this study is twofold. On the one hand, we find that the Eduscopio index can be considered as a good predictor of university performance of the students coming from a given school. On the other hand, our analysis shows that when the individual high schools' Eduscopio indexes are plotted along with their confidence intervals, most schools "overlap', meaning that is not possible to precisely rank them. This is not an unfamiliar result when building school-level (or even university-level) PIs, and stems from the limited number of students used to build the PIs, especially for small schools or schools with few

students enrolled in university.[18] However, the index is still informative to identify potentially problematic schools (i.e. schools at the bottom of the index distribution) or exceptionally performing schools (at the top of the distribution), especially at the school track-city level.

[18] Uncertainty (reflected in the confidence intervals) is a key element of any school league table (Leckie and Goldstein, 2011).

Table 1: Summary statistics at student's level (N=642,263)

Variables	(1) mean	(2) sd	(3) min	(4) max
Gender				
Male	0.432	0.495	0	1
Female	0.568	0.495	0	1
Age at matriculation				
Age	19.07	0.492	17	22
17	6.07e-05	0.00779	0	1
18	0.0638	0.244	0	1
19	0.823	0.381	0	1
20	0.0942	0.292	0	1
21	0.0159	0.125	0	1
22	0.00283	0.0532	0	1
Nationality				
Italian	0.976	0.152	0	1
Foreign	0.0209	0.143	0	1
Missing	0.00272	0.0521	0	1
High school track				
Classic Lyceum	0.169	0.375	0	1
Scientific Lyceum	0.452	0.498	0	1
Social Sciences Lyceum	0.0900	0.286	0	1
Linguistic Lyceum	0.0743	0.262	0	1
TI-Economics	0.115	0.319	0	1
TI-Technological	0.0995	0.299	0	1
High school final mark	79.15	12.04	60	101
Linear distance (km from school to university)	86.30	180.6	0	1,170
School Region				
Piedmont	0.0606	0.239	0	1
Valle d'Aosta	0.00104	0.0323	0	1
Lombardy	0.132	0.338	0	1
Trentino-Alto Adige	0.00607	0.0777	0	1
Veneto	0.0733	0.261	0	1
Friuli-Venezia Giulia	0.0174	0.131	0	1
Liguria	0.0224	0.148	0	1
Emilia-Romagna	0.0634	0.244	0	1
Tuscany	0.0548	0.228	0	1
Umbria	0.0146	0.120	0	1
Marche	0.0296	0.169	0	1
Lazio	0.0974	0.297	0	1
Abruzzo	0.0274	0.163	0	1
Molise	0.00722	0.0847	0	1
Campania	0.129	0.335	0	1
Apulia	0.0847	0.278	0	1
Basilicata	0.0147	0.120	0	1
Calabria	0.0460	0.209	0	1

Sicily	0.0926	0.290	0	1
Sardinia	0.0263	0.160	0	1
School year of Diploma				
2008/2009	0.329	0.470	0	1
2009/2010	0.346	0.476	0	1
2010/2011	0.325	0.468	0	1
Academic year of matriculation				
2009/2010	0.286	0.452	0	1
2010/2011	0.339	0.473	0	1
2011/2012	0.375	0.484	0	1
Years between diploma and matriculation				
0	0.911	0.285	0	1
1	0.0846	0.278	0	1
2	0.00422	0.0648	0	1
College major of matriculation				
Economics-Statistics	0.160	0.367	0	1
Law-Political Sciences	0.149	0.356	0	1
Medicine and Surgery	0.0407	0.198	0	1
Healthcare professions	0.0637	0.244	0	1
Mathematics, Physics and Natural Sciences	0.157	0.364	0	1
Science of physical and sport activities	0.0187	0.135	0	1
Social	0.0731	0.260	0	1
Technical	0.178	0.383	0	1
Humanistic	0.159	0.366	0	1

Table 2: Summary statistics at high school level (N=4,280)

Variables	(1) mean	(2) sd	(3) min	(4) max
Type of high school				
Classic Lyceum	0.1513	0.3584	0	1
Scientific Lyceum	0.3049	0.4604	0	1
Social Sciences Lyceum	0.1102	0.3132	0	1
Linguistic Lyceum	0.1322	0.3387	0	1
TI-Economics	0.1620	0.3685	0	1
TI-Tecnological	0.1394	0.3464	0	1
Region				
Piedmont	0.0665	0.249	0	1
Valle d'Aosta	0.00304	0.550	0	1
Lombardy	0.145	0.352	0	1
Trentino-Alto Adige	0.0105	0.102	0	1
Veneto	0.0707	0.256	0	1
Friuli-Venezia Giulia	0.0191	0.137	0	1
Liguria	0.0224	0.148	0	1
Emilia-Romagna	0.0630	0.243	0	1
Tuscany	0.0626	0.242	0	1
Umbria	0.0152	0.122	0	1
Marche	0.0311	0.173	0	1
Lazio	0.0983	0.298	0	1
Abruzzo	0.0276	0.164	0	1
Molise	0.00911	0.0950	0	1
Campania	0.102	0.303	0	1
Apulia	0.0703	0.256	0	1
Basilicata	0.0168	0.129	0	1
Calabria	0.0430	0.203	0	1
Sicily	0.0880	0.283	0	1
Sardinia	0.0355	0.185	0	1
Other characteristics				
Private school	0.130	0.336	0	1
Number of 13th graders	196.2	158.4	13	1,092
Number of university students	144.9	140.9	11	1,013
% of university students	71.68	20.34	30.87	100
% of females	57.53	25.13	0	100
% of foreign students	2.54	3.31	0	45.33

Table 3: Step 1. OLS estimations for first-year outcome variables

	First academic year					
	PUC			GPA		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Model1	Model2	Model3	Model1	Model2	Model3
Female	0.037***	0.016***	0.016***	0.473***	0.078***	0.079***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.007)	(0.007)
Foreign student	-0.021***	-0.013***	-0.012***	-0.305***	-0.133***	-0.115***
	(0.003)	(0.003)	(0.003)	(0.026)	(0.024)	(0.024)
Foreign missing	-0.032***	-0.022***	-0.023***	-0.268***	-0.219***	-0.218***
	(0.008)	(0.008)	(0.008)	(0.072)	(0.067)	(0.067)
Age	-0.047***	-0.050***	-0.049***	-0.028***	-0.132***	-0.131***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.008)	(0.008)
High school final mark	0.010***	0.010***	0.010***	0.090***	0.097***	0.098***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Private school	-0.108***	-0.116***	-0.117***	-0.559***	-0.625***	-0.636***
	(0.002)	(0.002)	(0.002)	(0.019)	(0.017)	(0.017)
One year delay	0.014***	-0.011***	-0.010***	0.511***	0.166***	0.167***
	(0.002)	(0.002)	(0.002)	(0.013)	(0.012)	(0.012)
Two years delay	0.018**	-0.017**	-0.016**	0.940***	0.509***	0.506***
	(0.008)	(0.007)	(0.007)	(0.059)	(0.054)	(0.054)
Linear distance (100 of km)	0.022***	0.008***	0.008***	0.016**	-0.038***	-0.034***
	(0.001)	(0.001)	(0.001)	(0.006)	(0.007)	(0.007)
Linear distance squared (100 of km)	-0.002***	-0.001***	-0.001***	-0.013***	0.003***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Enrolled in 2010	0.024***	0.024***	0.024***	-0.031***	-0.032***	-0.032***
	(0.001)	(0.001)	(0.001)	(0.009)	(0.008)	(0.008)
Enrolled in 2011	0.003**	0.006***	0.006***	-0.093***	-0.106***	-0.105***
	(0.001)	(0.001)	(0.001)	(0.009)	(0.008)	(0.008)
Constant	0.786***	0.843***	0.837***	18.667***	18.991***	19.050***
	(0.019)	(0.021)	(0.021)	(0.165)	(0.164)	(0.167)
Observations	642,214	642,214	642,214	538,701	538,701	538,701
R-squared	0.168	0.243	0.245	0.211	0.351	0.353
High school FE	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES
High school & Region	NO	NO	YES	NO	NO	YES
College major & University	NO	YES	YES	NO	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note. PUC and GPA stand for Percentage of University Credits and Grade Point Average, respectively.

Table 4: Step 1. OLS estimations for third-year outcome variables

	Third academic year					
		PUC			GPA	
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Model1	Model2	Model3	Model1	Model2	Model3
Female	0.051***	0.026***	0.026***	0.528***	0.143***	0.145***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.006)	(0.006)
Foreign student	-0.026***	-0.016***	-0.014***	-0.449***	-0.275***	-0.255***
	(0.003)	(0.003)	(0.003)	(0.023)	(0.021)	(0.021)
Foreign missing	-0.043***	-0.031***	-0.032***	-0.296***	-0.234***	-0.233***
	(0.008)	(0.007)	(0.007)	(0.063)	(0.057)	(0.057)
Age	-0.054***	-0.058***	-0.057***	-0.086***	-0.181***	-0.181***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)
High school final mark	0.009***	0.010***	0.010***	0.088***	0.095***	0.095***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Private school	-0.104***	-0.112***	-0.113***	-0.562***	-0.602***	-0.615***
	(0.002)	(0.002)	(0.002)	(0.016)	(0.015)	(0.015)
One year delay	0.007***	-0.027***	-0.027***	0.484***	0.126***	0.128***
	(0.002)	(0.002)	(0.002)	(0.012)	(0.011)	(0.011)
Two years delay	0.011***	-0.026***	-0.026***	0.657***	0.236***	0.238***
	(0.004)	(0.004)	(0.004)	(0.030)	(0.027)	(0.027)
Linear distance (100 of km)	0.028***	0.012***	0.013***	0.007	-0.046***	-0.042***
	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)	(0.006)
Linear distance squared (100 of km)	-0.002***	-0.001***	-0.001***	-0.012***	0.003***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Enrolled in 2010	0.007***	0.007***	0.007***	-0.043***	-0.045***	-0.044***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.007)	(0.007)
Enrolled in 2011	-0.029***	-0.024***	-0.024***	-0.096***	-0.105***	-0.104***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.007)	(0.007)
Constant	0.962***	1.048***	1.052***	20.077***	20.266***	20.364***
	(0.019)	(0.020)	(0.020)	(0.145)	(0.142)	(0.144)
Observations	642,214	642,214	642,214	558,858	558,858	558,858
R-squared	0.190	0.272	0.274	0.251	0.399	0.400
High school FE	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES
High school & Region	NO	NO	YES	NO	NO	YES
College major & University	NO	YES	YES	NO	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note. PUC and GPA stand for Percentage of University Credits and Grade Point Average, respectively.

Table 5: Step 1. Linear probability model for on-time graduation

	On-time graduation		
	(1)	(2)	(3)
Variables	Model1	Model2	Model3
Female	0.046***	0.015***	0.016***
	(0.001)	(0.001)	(0.001)
Foreign student	-0.037***	-0.026***	-0.021***
	(0.004)	(0.004)	(0.004)
Foreign missing	-0.035***	-0.023***	-0.022**
	(0.009)	(0.009)	(0.009)
Age	-0.042***	-0.042***	-0.042***
	(0.001)	(0.001)	(0.001)
High school final mark	0.010***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)
Private school	-0.077***	-0.095***	-0.099***
	(0.003)	(0.003)	(0.003)
One year delay	0.033***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)
Two years delay	0.041***	-0.001	-0.001
	(0.005)	(0.005)	(0.005)
Linear distance (100 of km)	0.040***	0.018***	0.019***
	(0.001)	(0.001)	(0.001)
Linear distance squared (100 of km)	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
Enrolled in 2010	0.021***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)
Enrolled in 2011	0.028***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)
Constant	0.349***	0.343***	0.371***
	(0.021)	(0.023)	(0.024)
Observations	587,546	587,544	587,544
R-squared	0.157	0.245	0.246
High school FE	YES	YES	YES
Region FE	YES	YES	YES
High school & Region	NO	NO	YES
College major & University	NO	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Summary statistics of raw and standardized (std) outcome variables

Variables	(1) N	(2) mean	(3) sd	(4) min	(5) max
Panel A					
PUC first-year	642,263	0.60	0.38	0	1
PUC std - first-year	642,214	0.60	0.37	-0.30	1.45
GPA first-year	538,742	24.93	2.80	18	31
GPA std - first-year	538,701	24.93	2.61	14.44	33.32
PUC third-year	642,263	0.57	0.37	0	1
PUC std - third-year	642,214	0.57	0.35	-0.35	1.35
GPA third-year	558,899	25.17	2.56	18	31
GPA std - third-year	558,858	25.17	2.37	15.32	32.82
Degree	587,581	0.31	0.46	0	1
Degree std	587,544	0.31	0.43	-0.58	1.41
Panel B					
PUC index first-year	4,280	55.54	13.49	0	100
GPA index first-year	4,280	64.52	11.96	0	100
ES index first-year	4,280	60.03	11.95	1.49	94.08
PUC index third-year	4,280	55.50	13.75	0	100
GPA index third-year	4,280	61.48	13.21	0	100
ES index third-year	4,280	58.49	12.85	2.43	92.16
DEGREE index	4,280	42.64	12.89	0	100

Note. *Panel A* reports the raw summary statistics and *Panel B* the summary statistics of the residuals, after raw performance measures have been purged from college major-Higher Education Institution fixed effects. PUC, GPA and ES stand for Percentage of University Credits, Grade Point Average and Eduscopio, respectively.

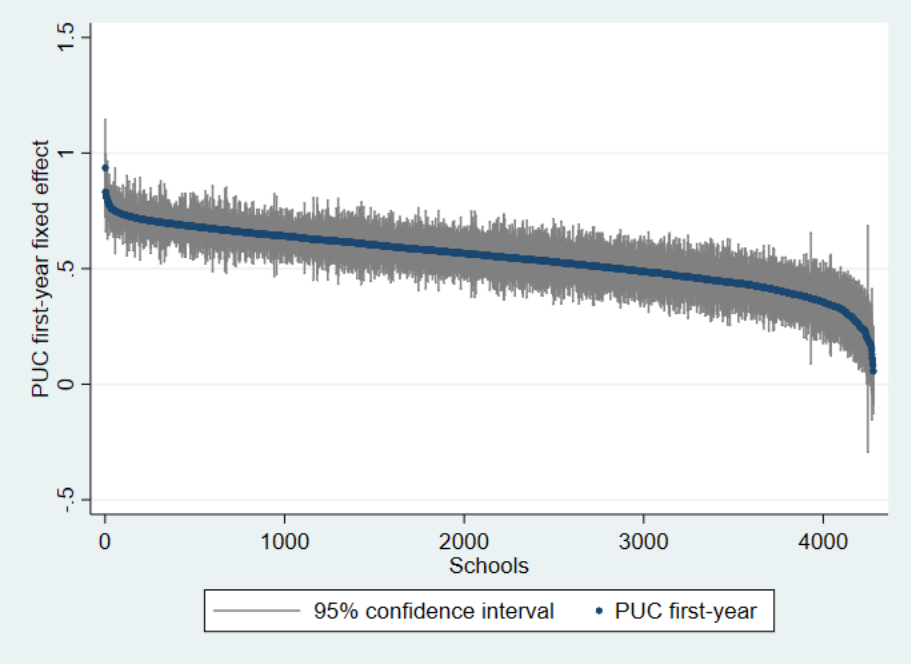
Table 7: Linear probability model of ranking decile transition matrix

Variables	(1) Main Diagonal	(2) Above Diagonal	(3) Below Diagonal
Scientific Lyceum	-0.003	-0.058***	0.061***
	(0.022)	(0.019)	(0.015)
Social Sciences Lyceum	-0.093***	0.056**	0.037*
	(0.029)	(0.026)	(0.020)
Linguistic Lyceum	-0.104***	0.032	0.071***
	(0.027)	(0.024)	(0.020)
TI-Economics	-0.095***	-0.071***	0.165***
	(0.026)	(0.021)	(0.020)
TI-Technological	-0.130***	-0.046**	0.176***
	(0.027)	(0.022)	(0.022)
Private	-0.045**	0.035*	0.010
	(0.023)	(0.019)	(0.018)
Constant	0.719***	0.178***	0.103***
	(0.033)	(0.026)	(0.025)
Observations	4,280	4,280	4,280
R-squared	0.018	0.023	0.035
Region FE	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

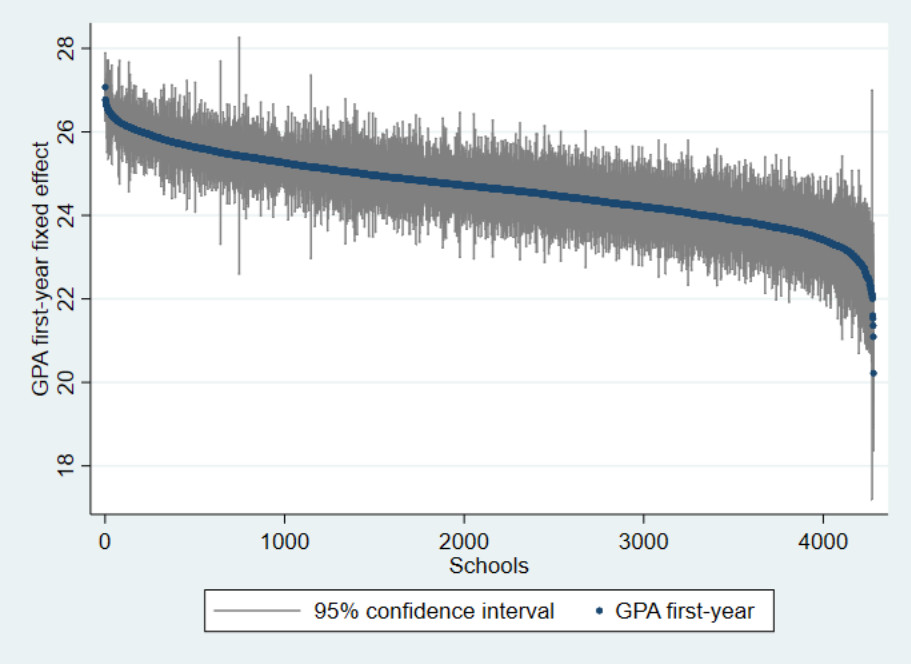
Note. Probability of remaining on the main diagonal (column 1), moving above the main diagonal (column 2) or below the main diagonal (column 3) in the school ranking deciles transition matrix (from first-year to three-year performance).

Figure 1: School ranking by first-year PUC



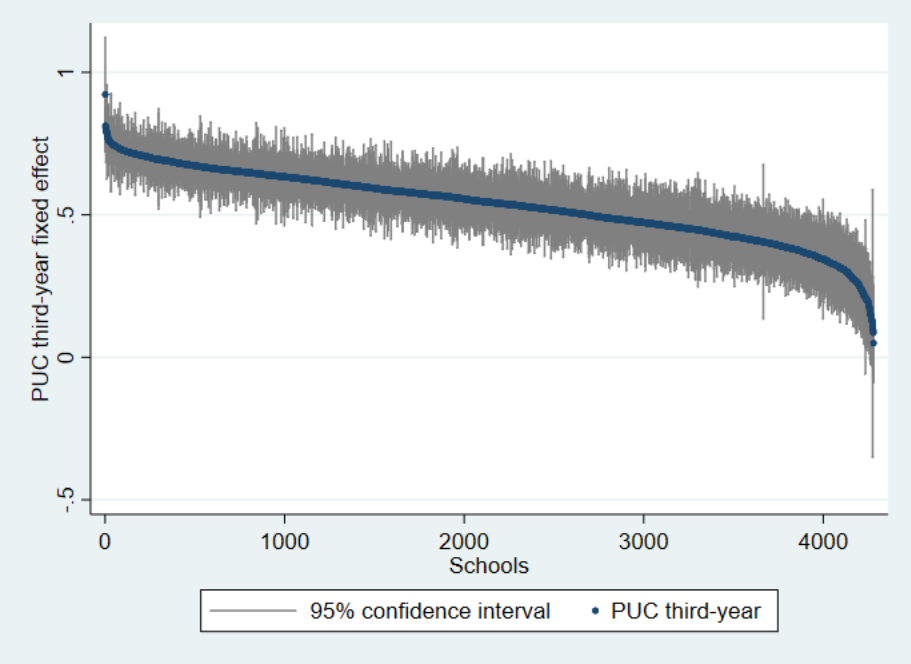
Note. This figure plots the high school fixed effects and 95% confidence intervals for first-year Percentage of University Credits (PUC).

Figure 2: School ranking by first-year GPA



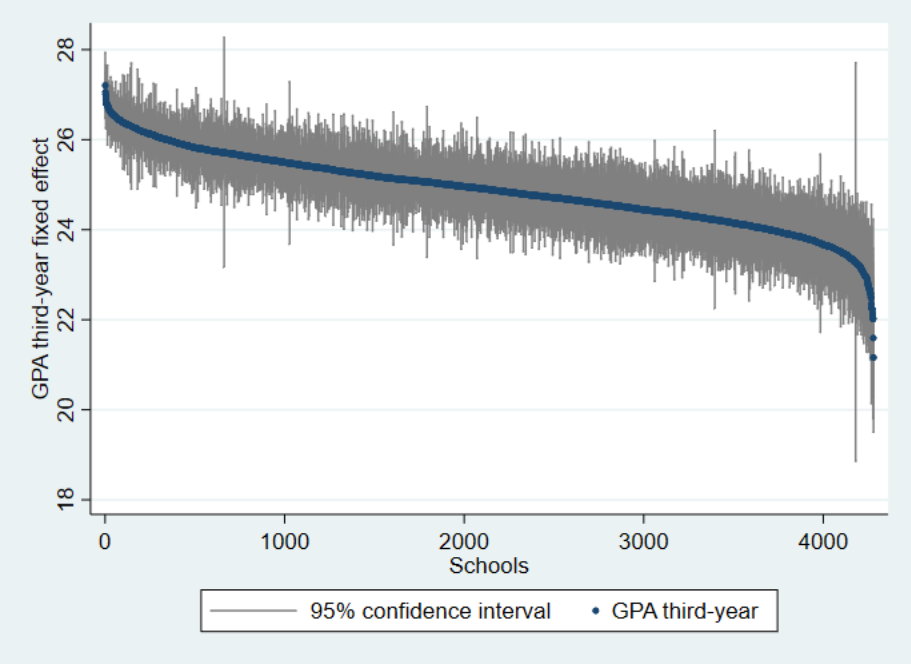
Note. This figure plots the high school fixed effects and 95% confidence intervals for first-year Grade Point Average (GPA).

Figure 3: School ranking by third-year PUC



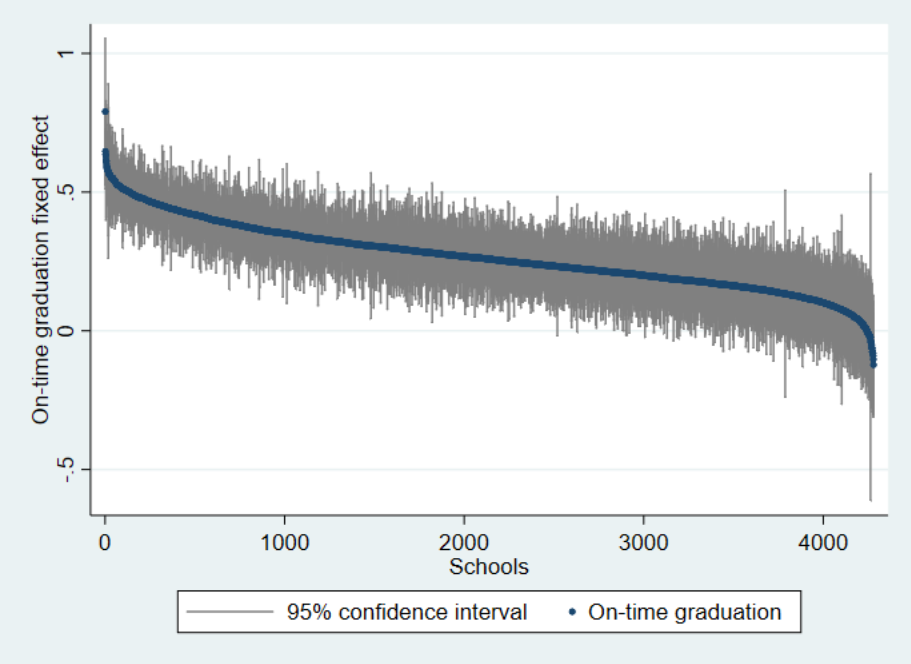
Note. This figure plots the high school fixed effects and 95% confidence intervals for third-year Percentage of University Credits (PUC).

Figure 4: School ranking by third-year GPA



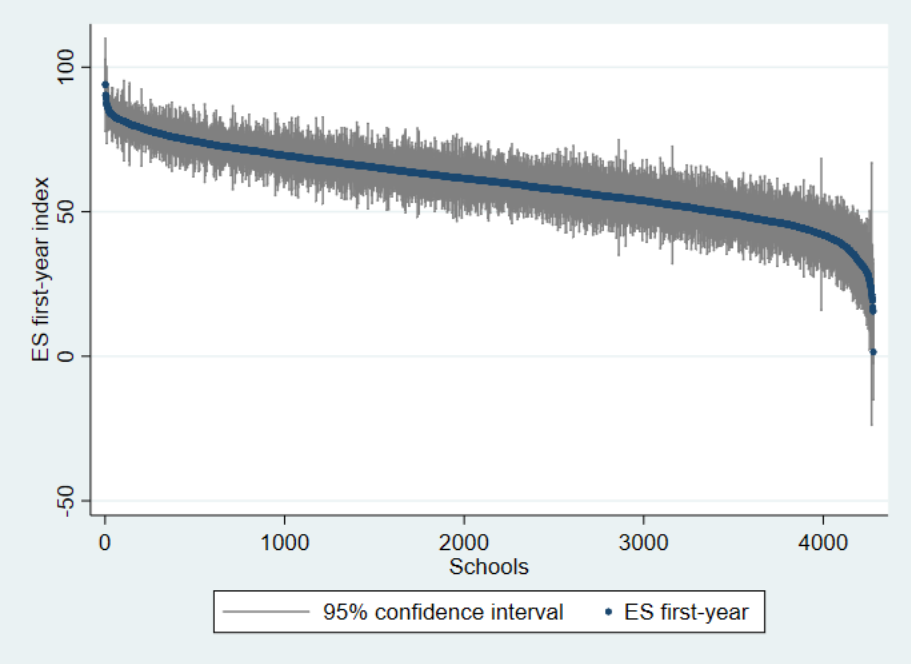
Note. This figure plots the high school fixed effects and 95% confidence intervals for third-year Grade Point Average (GPA).

Figure 5: School ranking by on-time graduation fixed effect



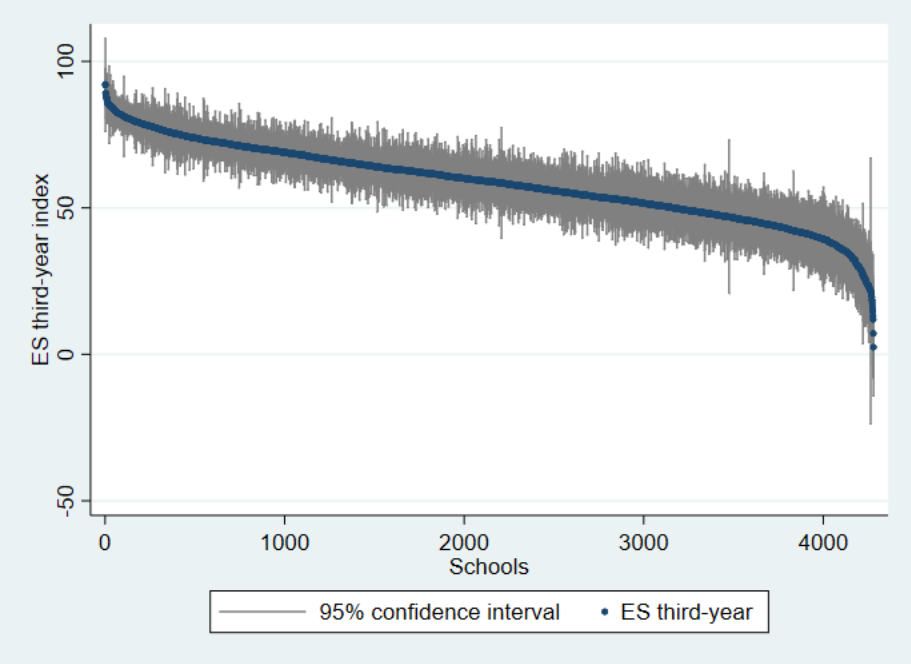
Note. This figure plots the high school fixed effects and 95% confidence intervals for on-time graduation.

Figure 6: School ranking by first-year Eduscopio (ES) index



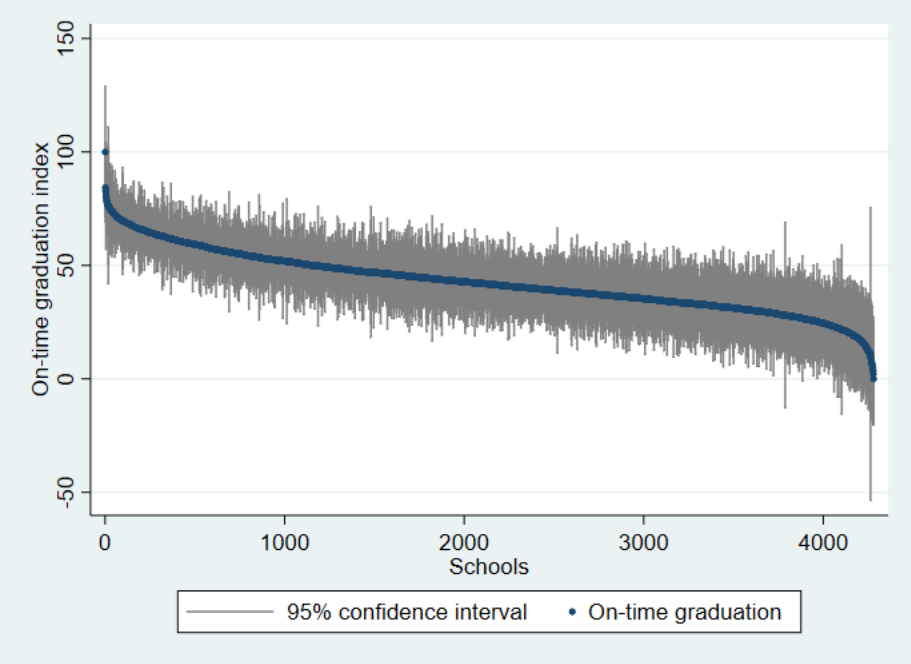
Note. This figure plots the values of the first-year Eduscopio index with 95% confidence intervals.

Figure 7: School ranking by third-year Eduscopio (ES) index



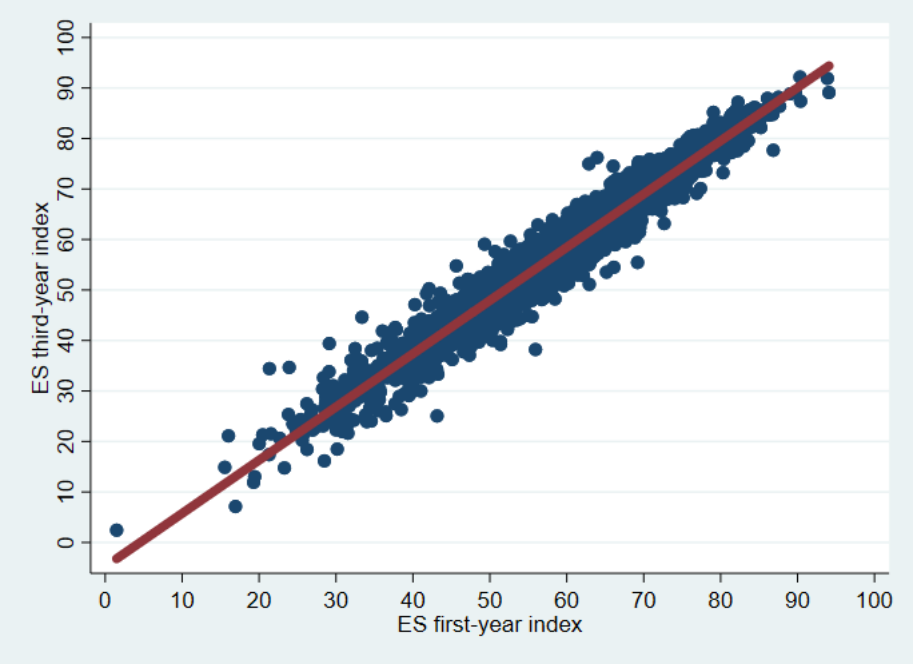
Note. This figure plots the values of the third-year Eduscopio index with 95% confidence intervals.

Figure 8: School ranking by on-time graduation index



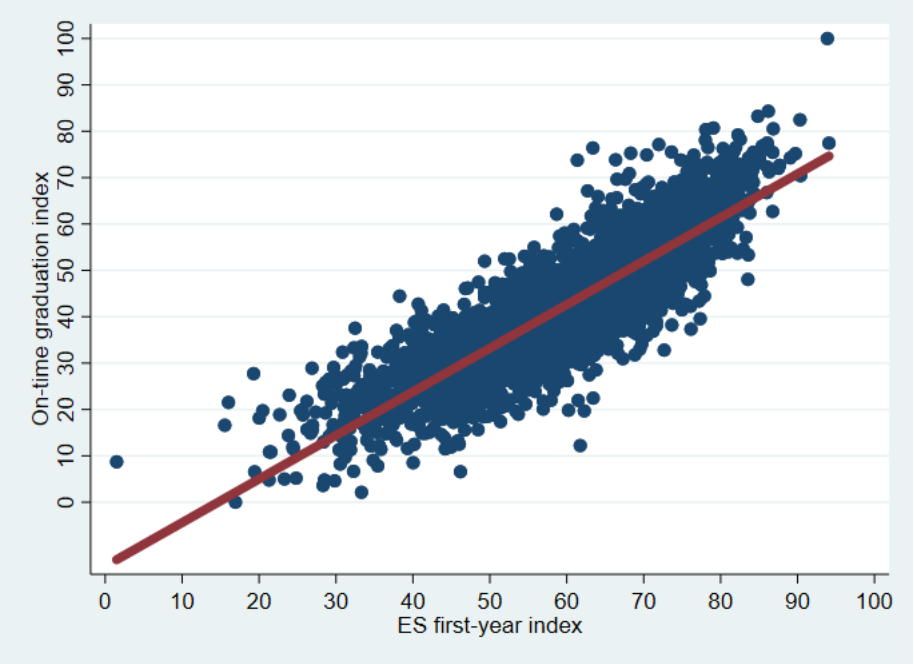
Note. This figure plots the values of on-time graduation index with 95% confidence intervals.

Figure 9: Correlation between first- and third-year Eduscopio (ES) index



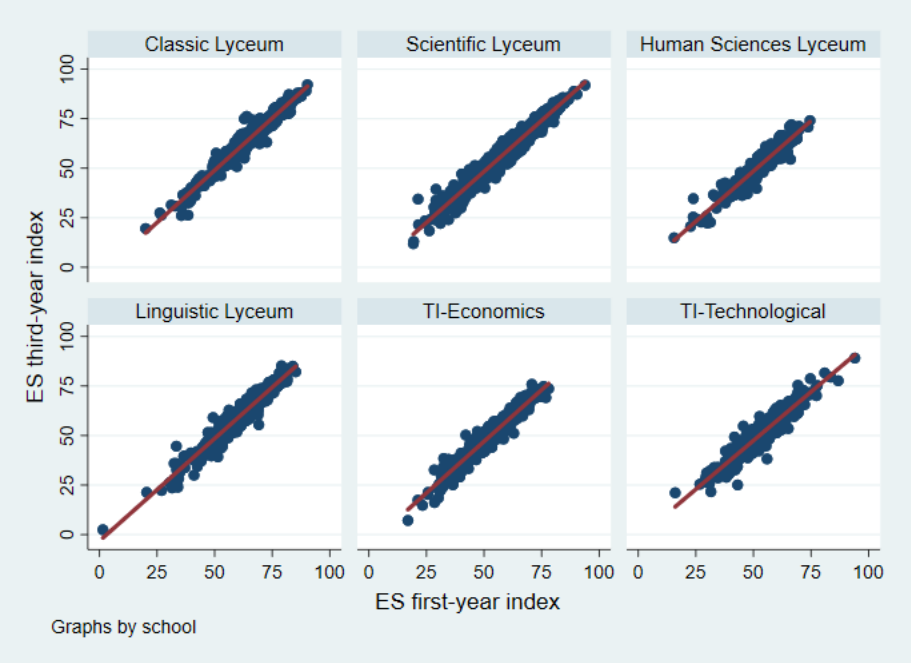
Note. This figure reports the cross-plot between the first- and the third-year Eduscopio index. The equation of the regression line is $ES3_i = -4.78(0.26) + 1.05(0.004) * ES1_i$ (standard errors in parentheses).

Figure 10: Correlation between on-time graduation and first-year Eduscopio (ES) index



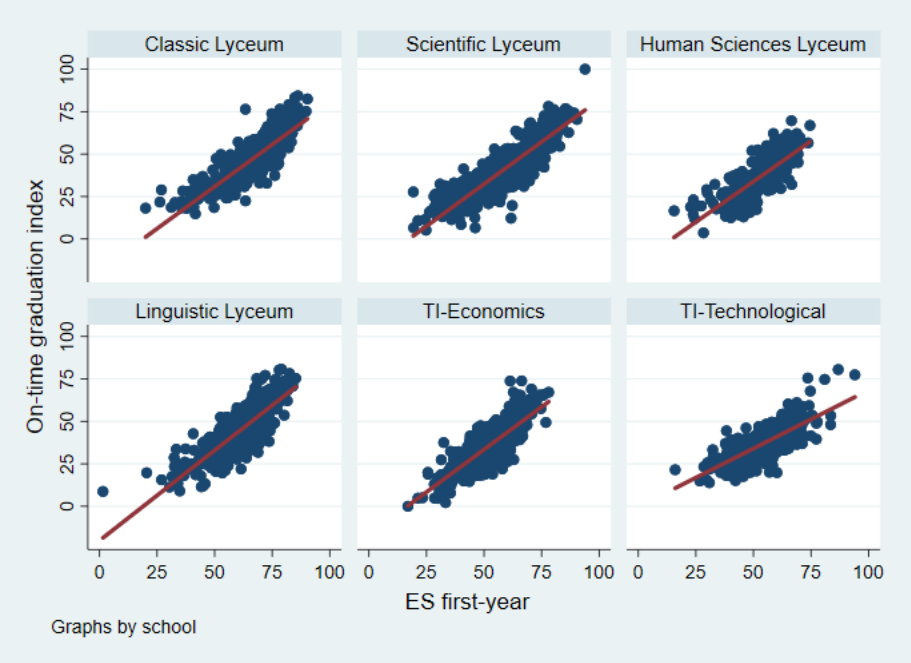
Note. This figure reports the cross-plot between on-time graduation and the first-year Eduscopio index. The equation of the regression line is $DEGREE_i = -13.80(0.56) + 0.94(0.009) * ES1_i$ (standard errors in parentheses).

Figure 11: Correlation between first- and third-year Eduscopio (ES) index by school type



Note. This figure reports the cross-plot between the first- and the third-year Eduscopio index by school type. The βs for the different school tracks are: $\beta(\textit{Classic Lyceum}) = 1.05(0.01)$, $\beta(\textit{Scientific Lyceum}) = 1.03(0.01)$, $\beta(\textit{Human Sciences Lyceum}) = 1.01(0.02)$, $\beta(\textit{Linguistic Lyceum}) = 1.03(0.01)$, $\beta(\textit{TI-Economics}) = 1.04(0.01)$ and $\beta(\textit{TI-Technological}) = 0.98(0.01)$ (standard errors in parentheses).

Figure 12: Correlation between on-time graduation and the first-year Eduscopio (ES) index by school type



Note. This figure reports the cross-plot between the on-time graduation and the first-year Eduscopio index by school type. The βs for the different school tracks are: $\beta(\textit{Classic Lyceum}) = 0.99(0.03)$, $\beta(\textit{Scientific Lyceum}) = 0.99(0.02)$, $\beta(\textit{Human Sciences Lyceum}) = 0.95(0.04)$, $\beta(\textit{Linguistic Lyceum}) = 1.06(0.04)$, $\beta(\textit{TI-Economics}) = 1.002(0.03)$ and $\beta(\textit{TI-Technological}) = 0.69(003)$ (standard errors in parentheses).

References

Aina, C., E. Baici, and G. Casalone (2011). Time to degree: students' abilities, university characteristics or something else? Evidence from Italy. *Education Economics 19*(3), 311–325.

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2008). Differential grading standards and university funding: Evidence from Italy. *CESifo Economic Studies 54*(2), 149–176.

Bernardi, M. and G. De Simone (2018). eduscopio.it Esiti Universitari. Technical report, Documento tecnico, Fondazione Agnelli. Torino.

Bertola, G., D. Checchi, and V. Oppedisano (2007). Private school quality in Italy. *Giornale degli Economisti e Annali di Economia 66*(3), 375–400.

Card, D. and A. A. Payne (2017). High school choices and the gender gap in stem. Technical report, National Bureau of Economic Research.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *Review of Economics and Statistics 94*(3), 699–711.

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review 29*(6), 911–922.

Jamelske, E. (2009). Measuring the impact of a university first-year experience program on student GPA and retention. *Higher Education 57*(3), 373–391.

Kahn, S. and D. Ginther (2017). Women and STEM. Technical report, National Bureau of Economic Research.

Lang, D. (2007). The impact of a first-year experience course on the academic performance, persistence, and graduation rates of first-semester college students at a public research university. *Journal of the First-Year Experience & Students in Transition 19*(1), 9–25.

Leckie, G. and H. Goldstein (2011). Understanding uncertainty in school league tables. *Fiscal Studies 32*(2), 207–224.

Porter, S. R. and R. L. Swing (2006). Understanding how first-year seminars affect persistence. *Research in Higher Education 47*(1), 89–109.

Riegle-Crumb, C., B. King, E. Grodsky, and C. Muller (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal 49*(6), 1048–1073.

Vuri, D. (2018). Ranking scolastici e scelte familiari. Prime evidenze da Eduscopio. Technical report, Working Paper 58 Fondazione Agnelli, Fondazione Agnelli. Torino.

A Appendix: Confidence intervals

In this section we describe the steps follwed to build the Eduscopio indexes' confidence intervals. We follow the same procedure described in the Eduscopio technical reports (Bernardi and De Simone, 2018) :

Step 1 We estimate the school fixed effects for the GPAand PUC, θ_1^i and θ_2^i, respectively

Step 2 Then, we normalize the fixed effects as follows:

$$\theta_j^{norm,i} = \frac{\theta_j^i - \theta_j^{min}}{\theta_j^{max} - \theta_j^{min}} \tag{9}$$

where $j = 1, 2$ is the outcome subscript and i is the school subscript. For easy of notation $A_j := \theta_j^{max} - \theta_j^{min}$.

Step 3 We build the Eduscopio index

$$ES^i = \frac{1}{2} \cdot \theta_1^{norm,i} + \frac{1}{2} \cdot \theta_2^{norm,i} \tag{10}$$

Step 4 We build the variance of the Eduscopio index

$$var(ES^i) = \frac{1}{4} \cdot \frac{var(\theta_1^i)}{(\theta_1^{max} - \theta_1^{min})^2} + \frac{1}{4} \cdot \frac{var(\theta_2^i)}{(\theta_2^{max} - \theta_2^{min})^2} + \frac{1}{2} \cdot \frac{cov(\theta_1^i, \theta_2^i)}{(\theta_1^{max} - \theta_1^{min}) \cdot (\theta_2^{max} - \theta_2^{min})} \tag{11}$$

where $var(\theta_j^i) = (se[\hat{\theta_j^i}])^2$ (i.e. the square of the school FEs).

$cov(\theta_1^i, \theta_2^i)$ is replaced with the sample analogue $cov(\hat{\theta_1^i}, \hat{\theta_2^i})$

Step 4 We compute the 95% confidence interval:

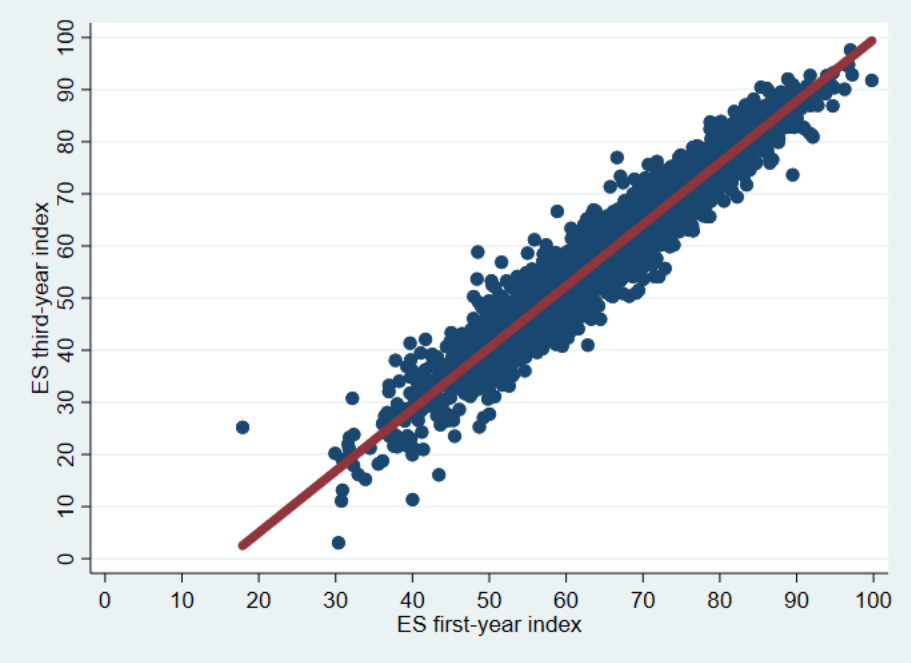
$$ES_i \pm 1.96 \cdot \frac{\sqrt{var(ES^i)}}{\sqrt{n_i}}$$

where n_i is the number of students from school i.

B Appendix: Robustness checks

In this section we investigate the robustness of the ES indicators using the 31st October as the end of the selected academic year (first and third). We do this to minimize the risk of including in our PIs also the exams achieved but not belonging to the first or to the third academic year, as we do not have precise information on how the exams are distributed across different academic year. By implementing this new threshold, Figure B1 shows that the correlation between the first and the third-year Eduscopio indexes is strong ($\beta = 1.18$) and statistically significant. This result is in line with Figure 9 ($\beta = 1.05$).

Figure B1: Correlation between first- and third-year Eduscopio (ES) index (academic year ends at 31st October)



Note. This figure reports the cross-plot between the first- and the third-year Eduscopio index when the end of the academic year is set at 31st October. The equation of the regression line is $ES3(31OCT)_i = -18.66(0.41) + 1.18(0.006) * ES1(31OCT)_i$ (standard errors in parentheses).
The other correlations between the indexes (with the academic year ending at 31st October) not shown in the figure are:
$ES3_i = -4.78(0.26) + 1.05(0.004) * ES1_i$
$DEGREE_i = -13.80(0.56) + 0.94(0.009) * ES1_i$
$ES3\_100_i = 0.96(0.15) + 0.98(0.002) * ES1\_100_i$
$DEGREE\_100_i = 6.06(0.37) + 0.88(0.006) * ES1\_100_i$
where variables with 100 indicate percentiles.

C Appendix: Rankings by city

Table C1: School ranking: academic track (Rome)

	First year			Third year		
	(1)	(2)	(3)	(4)	(5)	(6)
School name	Rank	ES index	CI	Rank	ES index	CI
Top1	1	83.8	81.4-86.1	1	84.8	82.5-87.1
Top2	2	82.3	76.9-87.7	6	81.1	75.8-86.4
Top3	3	82.2	78.9-85.4	2	84.6	81.4-87.8
Top4	4	82.0	79.1-85.0	3	83.2	80.3-86.1
Top5	5	81.5	78.0-84.9	4	82.2	78.8-85.6
Top6	6	79.9	77.3-82.5	7	80.7	78.1-83.3
Top7	7	79.6	76.6-82.6	5	81.2	78.2-84.2
Top8	8	78.2	75.7-80.7	8	79.9	77.4-82.4
Top9	9	77.3	74.8-79.8	13	77.1	74.7-79.5
Top10	10	77.2	73.2-81.1	12	77.3	73.4-81.3
Bottom1	115	35.4	27.2-43.6	116	32.1	24.1-40.1
Bottom2	116	35.4	25.1-45.6	108	38.5	28.2-48.7
Bottom3	117	33.8	26.2-41.4	118	30.7	23.3-38.1
Bottom4	118	32.7	22.0-43.4	115	34.3	23.8-44.8
Bottom5	119	31.9	22.6-41.2	120	27.7	18.4-37.0
Bottom6	120	30.6	22.4-38.9	122	27.6	19.7-35.4
Bottom7	121	29.4	22.5-36.4	119	29.9	23.2-36.6
Bottom8	122	28.6	19.1-38.2	121	27.7	18.1-37.2
Bottom9	123	26.2	2.2-50.3	123	27.5	3.5-51.5
Bottom10	124	24.4	14.1-34.8	124	23.5	13.3-33.7

Note. This table shows the first 10 and last 10 academic track schools in Rome according to the first-year Eduscopio index, along with the rank, the value of the index and the confidence interval, in column (1), (2) and (3), respectively. Columns (4), (5) and (6) report the corresponding values for the third-year Eduscopio index.

Table C2: School ranking: Human Sciences and Linguistic track (Rome)

	First year			Third year		
	(1)	(2)	(3)	(4)	(5)	(6)
School name	Rank	ES index	CI	Rank	ES index	CI
Top1	1	71.2	67.5-75.0	1	71.4	67.7-75.1
Top2	2	69.2	61.9-76.5	5	65.0	57.8-72.2
Top3	3	67.4	58.9-75.9	2	66.5	58.0-74.9
Top4	4	66.2	58.7-73.6	8	64.0	56.6-71.5
Top5	5	66.0	56.6-75.3	9	63.5	54.3-72.7
Top6	6	65.2	60.8-69.7	6	64.6	60.2-69.0
Top7	7	65.2	60.2-70.1	4	65.7	60.8-70.6
Top8	8	64.8	60.4-69.2	7	64.1	59.8-68.4
Top9	9	62.5	54.7-70.3	12	60.8	53.1-68.4
Top10	10	62.0	57.7-66.4	11	61.9	57.7-66.2
Bottom1	40	45.6	41.0-50.1	41	41.3	36.9-45.8
Bottom2	41	44.4	39.7-49.2	39	41.9	37.2-46.6
Bottom3	42	44.2	27.5-60.9	43	37.1	20.4-53.8
Bottom4	43	41.4	30.9-51.9	40	41.6	31.3-51.9
Bottom5	44	33.5	22.4-44.6	46	32.7	21.8-43.6
Bottom6	45	33.3	21.8-44.9	44	35.9	24.4-47.4
Bottom7	46	31.5	22.6-40.5	47	23.6	14.7-32.4
Bottom8	47	26.9	17.0-36.9	48	22.3	12.5-32.0
Bottom9	48	23.9	7.2-40.6	45	34.7	18.5-50.8
Bottom10	49	20.5	7.0-34.0	49	21.3	7.8-34.9

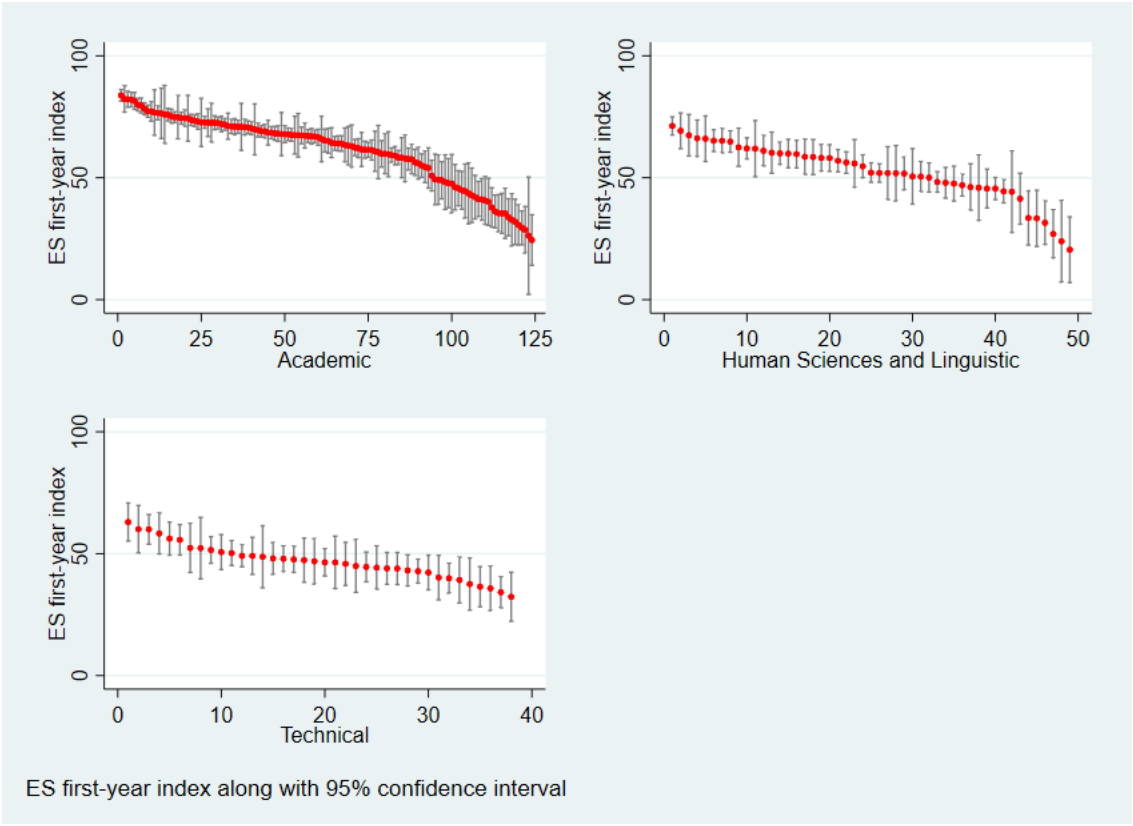
Note. This table shows the first 10 and last 10 Human sciences and Linguistic track schools in Rome according to the first-year Eduscopio index, along with the rank, the value of the index and the confidence interval, in column (1), (2) and (3), respectively. Columns (4), (5) and (6) report the corresponding values for the third-year Eduscopio index.

Table C3: School ranking: Technical track (Rome)

	First year			Third year		
	(1)	(2)	(3)	(4)	(5)	(6)
School name	Rank	ES index	CI	Rank	ES index	CI
Top1	1	63.0	55.2-70.8	2	57.5	49.8-65.1
Top2	2	60.1	50.3-69.8	3	56.6	47.3-66.0
Top3	3	60.0	53.9-66.0	1	58.6	52.7-64.6
Top4	4	58.3	49.9-66.7	5	52.8	44.5-61.1
Top5	5	56.2	49.4-62.9	4	53.8	47.2-60.5
Top6	6	55.7	49.4-61.9	6	51.3	45.2-57.3
Top7	7	52.3	42.3-62.4	10	47.4	37.5-57.3
Top8	8	52.2	39.6-64.9	9	47.8	35.2-60.4
Top9	9	51.5	46.1-57.0	11	46.4	41.0-51.9
Top10	10	50.7	43.5-57.8	7	48.8	41.7-55.9
Bottom1	29	42.8	37.9-47.7	31	38.4	33.5-43.2
Bottom2	30	42.3	35.2-49.4	32	38.2	31.2-45.2
Bottom3	31	40.2	31.1-49.3	22	43.5	34.4-52.6
Bottom4	32	39.9	33.8- 46.1	33	35.7	29.6-41.8
Bottom5	33	39.2	29.8-48.6	30	38.9	29.5-48.2
Bottom6	34	37.5	26.8-48.2	34	34.1	23.6-44.7
Bottom7	35	36.4	28.2-44.7	37	25.9	17.9-33.8
Bottom8	36	35.8	26.7-44.9	36	29.7	20.5-38.9
Bottom9	37	34.2	27.8-40.5	35	30.4	24.1-36.7
Bottom10	38	32.3	22.3-42.4	38	24.2	14.1-34.3

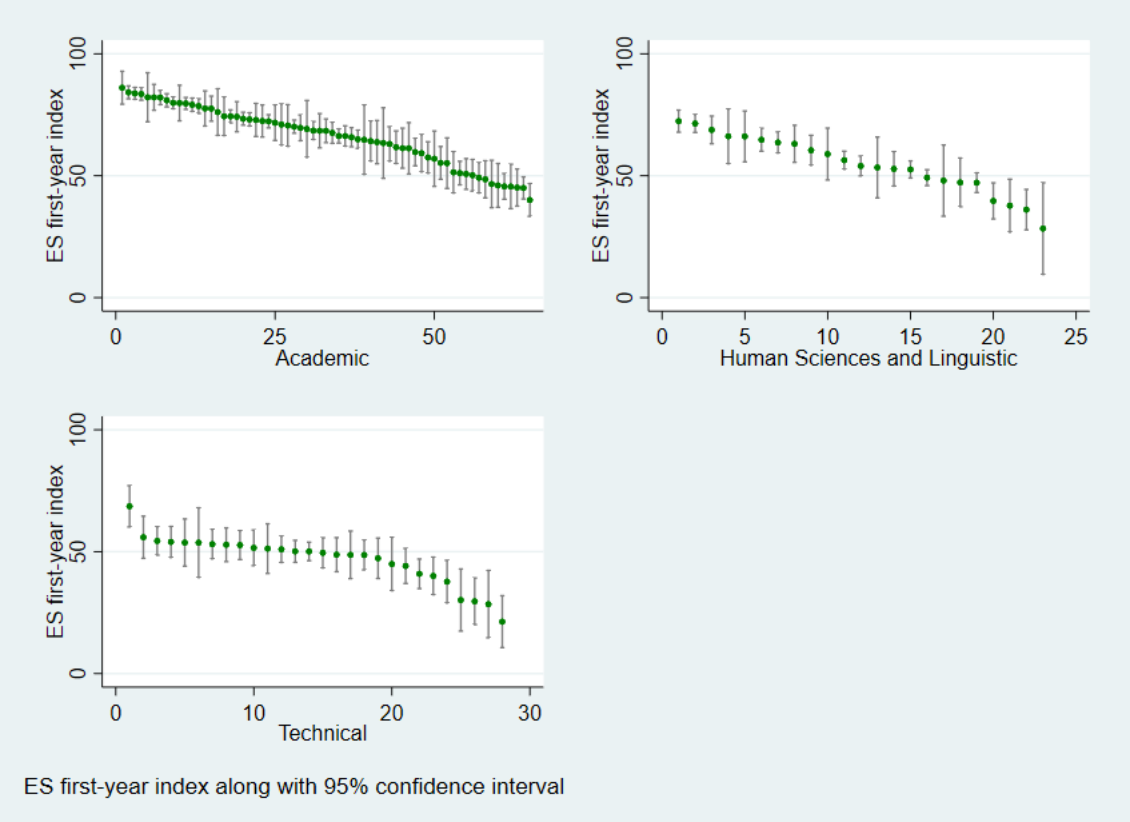
Note. This table shows the first 10 and last 10 Human technical track schools in Rome according to the first-year Eduscopio index, along with the rank, the value of the index and the confidence interval, in column (1), (2) and (3), respectively. Columns (4), (5) and (6) report the corresponding values for the third-year Eduscopio index.

Figure C1: Rome



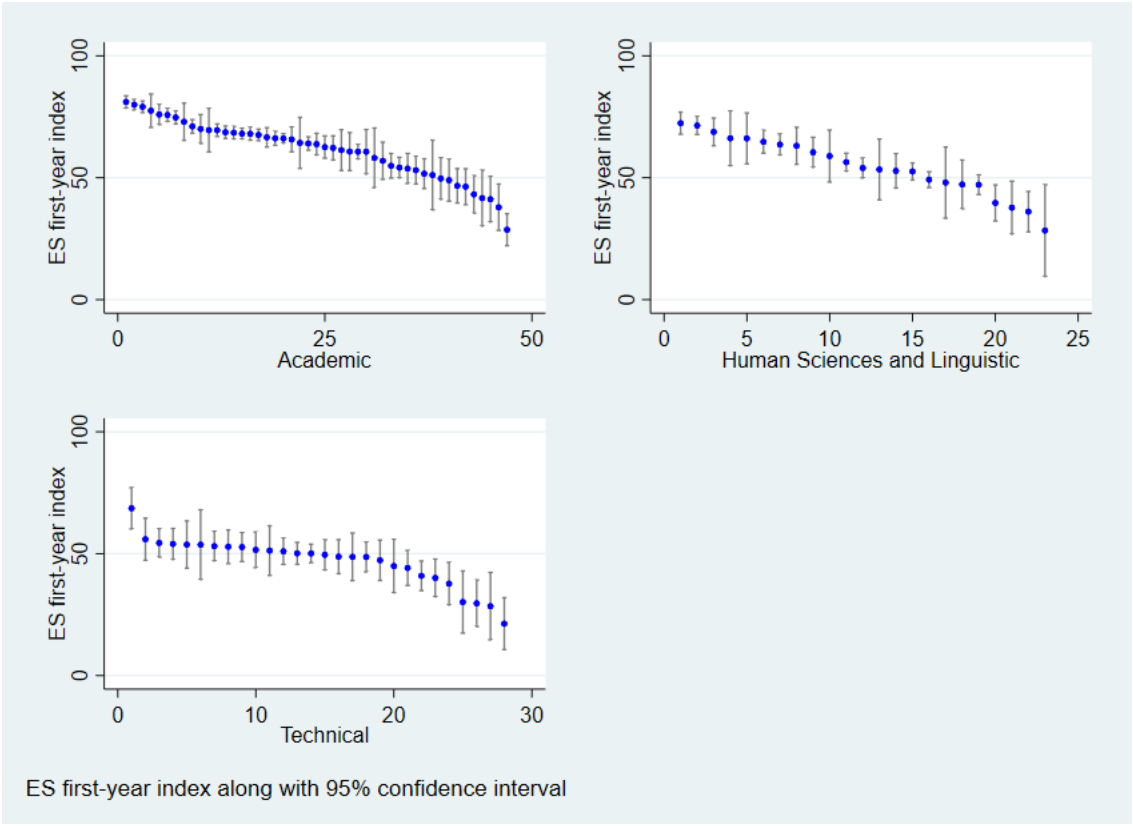
Note. First-year Eduscopio index and confidence interval by school track in the city of Rome.

Figure C2: Milan



Note. First-year Eduscopio index and confidence interval by school track in the city of Milan.

Figure C3: Naples



Note. First-year Eduscopio index and confidence interval by school track in the city of Naples.

n. 59 (04/2019)

The contribution of high schools to university students' academic performance: The case of Eduscopio

di Carmen Aina, Massimiliano Bratti, Enrico Lippo

Fondazione Agnelli