SIS | 2022

51st Scientific Meeting of the Italian Statistical Society

Caserta, 22-24 June

Book of the Short Papers

Editors: Antonio Balzanella, Matilde Bini, Carlo Cavicchia, Rosanna Verde

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*
Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cossari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Contents

4 Contributed Sessions 844

Does taking additional Maths classes improve university performance?

A clustered case-control study based on Italian high-school and university data

Martina Vittorietti, Andrea Priulla, Massimo Attanasio

Abstract Several recent studies in educational literature showed how students' skills in maths affect their success at higher levels of education. The aim of this paper is to evaluate the effect of taking additional maths class at high school on first-year performance of Italian university students. However, university performance and the choice of the high-school depend on several factors that make this evaluation challenging. Using information coming from three different sources, we carry out a multilevel propensity score procedure to estimate the average treatment effect between the applied sciences track and the traditional scientific one. After balancing for school- and student-level covariates, the results of a logistic regression model suggest no difference between the two school tracks.
Abstract Studi recenti nell'ambito dell'istruzione secondaria hanno mostrato come le abilità matematiche degli studenti abbiano un effetto sul loro futuro successo scolastico e universitario. L'obiettivo di questo lavoro è valutare se studiare più matematica a scuola è associato ad una migliore performance al primo anno di università. La performance universitaria e la scelta della scuola dipendono tuttavia da molteplici fattori che rendono complessa questa analisi. Utilizzando dati provenienti da tre diverse fonti, applichiamo un multilevel propensity score matching per stimare l'effetto medio di due trattamenti: "il liceo scientifico delle scienze applicate" e il "liceo scientifico tradizionale". Dopo aver bilanciato le caratteristiche della scuola e degli studenti, i risultati di un modello logistico mostrano che non emergono differenze tra i due percorsi scolastici.

Key words: propensity score, fine balance, educational data, multilevel propensity score, optimal network flow

Martina Vittorietti, Andrea Priulla, Massimo Attanasio
Department of Economics, Business and Statistics, University of Palermo, Italy, e-mail: martina.vittorietti@unipa.it, andrea.priulla@unipa.it, massimo.attanasio@unipa.it

1 Introduction

Mathematics is perceived as the foundation for scientific and technological knowledge that is cherished by societies worldwide. Due to its relevance and application, mathematics is widely regarded as one of the most important school subjects and a central major of the school curriculum [8]. mathematics underpins the study of many disciplines at university, not only in science, technology, engineering and mathematics (STEM) but also in agriculture, pharmacy and economics [10]. In Italy, there is at least one mathematics or mathematics related course in most of the degree courses, and it is often regarded as one of the toughest exams. Several surveys at both international level, such as the PISA tests promoted by OECD, and at national level, such as INVALSI tests, are conducted for investigating mathematical skills in high school. Italy is characterized by a high school system organized in separate tracks, which sorts students according to their academic or vocational orientation. Despite the existence of different tracks, the high school could be described as tripartite, with a 5-year academic oriented generalist education provided by lyceums, with further distinctions in humanities, science activities, languages, pedagogical sciences, 5-year technical schools, and 5-year vocational schools. In 2010, Italy underwent a reform of the secondary school education. One of the main innovations was the introduction of a new school curriculum: the applied sciences curriculum. The main difference with respect to the traditional scientific track was the substitution of the Latin language class with classes in scientific subjects such as mathematics, physics, natural sciences, and the addition of computer science as a new subject. In [12] the author compares the in-going and out-going mathematical skills of the students of the two different curricula, underlying the strengths and the weaknesses of both tracks, but as far as the authors' knowledge there are no results about the academic performances related to this new high school track. Therefore, the main objective of this paper is to compare the academic performances of the students coming from traditional scientific or applied sciences tracks, in order to answer the following question: "Does taking additional maths classes improve the university performance?" Before answering this question, it should be noted that university performance lies above an incredibly intricate net of multi-dimensional factors. Therefore, the question in the present form is ill-posed because of social and cultural factors that affect the choice of studying more mathematics or more correctly the choice of the applied sciences track, that eventually will affect university choice and the university performance too [1]. Among the social and cultural factors, the most relevant are: gender [3, 6], social level, location of the high school [2, 11], socio-economic status of the family [4, 16]. For investigating the difference between the two tracks, a balancing for all the possible factors related to both the school track and the academic performance is needed. Analogously to clinical studies, we investigate the effect of the *treatment* "Studying more maths". It is possible to see this study as a clustered case-control study, in which the cases are the students who attended the applied sciences track, and the control are the students who attended the traditional Scientific track. We propose a multi-level balancing approach of the the socio-demographic characteristics using a multi-level propensity score model. This procedure aims to eliminating the selection bias due to the fact that, in observational studies, the choice of the high school track, is not random, because it is related to socio-economic and cultural factors[4]. The number of ECTS achieved by the two group of students, treated and controls, in their first academic year is then compared and the average treatment effect (ATE) is estimated.

2 Data

The dataset used in this paper is built by linking three distinct administrative sources: aggregate data coming from the National Archive of Schools (ANS-S); micro-data on high school career from the National Evaluation Institute for the School System (INV);

As most of the educational data, the data at hand presents a hierarchical structure. In fact, it is possible to distinguish variables based on two levels:

1. *school-level*: macro-region, university enrollment rate and the Economic, Social, Cultural Status index (ESCS) used by INVALSI [14];
2. *student-level*: gender, ESCS, high school career's regularity.

3 Methods

Educational data often presents a multilevel structure, so, it is common to speak of clustered treatments. In clustered-treatment observational studies, treatments are applied to entire cluster (schools) or to individuals (students or teachers), or to multiple levels simultaneously (schools and students) [13]. Let S be the number of matched pairs of schools $s = 1,\dots,S$. Each school presents both classes belonging to the applied sciences and the traditional scientific track, $j = 1,2$, the treatment and the control respectively, for a total of $2S$ units. Each school track of each school s_j contains $n_{s_j} > 1$ students, $i = 1,\dots,n_{s_j}$. Each pair is matched on a vector of observed pre-treatment covariates $x_{s_{j_i}}$. Let X_{s_j} be the matrix whose rows consists of the $x_{s_{j_i}}$ vectors for each student i in the j-th track of the s-th school with support $\mathbf{X} \subset \mathbf{R}$. A student i in school track j of school s is described by both observed covariates and possibly unobserved covariates $u_{s_{j_i}}$. In our study, treatment assignment occurs at class level as whole classes are assigned to treatment or to control. If the j-th class in pair s receives the treatment $Z_{s_j} = 1$, whereas if it receives the control $Z_{s_j} = 0$, thus $Z_{s_1} + Z_{s_2} = 1$, for each s as each pair contains the treatment classes and the control classes. Following the designed structure, the statistical analysis strategy uses a propensity score model as a matching strategy [9, 15]. In particular, we follow the optimal multilevel matching using network flows proposed in [13], in which first we match the schools on school-related covariates and then we match students using individual socio-demographic characteristics. After the matching procedure, we fit a logit regression model with random effects on the matched data, to evaluate the effect of the treatment on the probability of achieving at least 37 ECTS. We chose the cut-point of 37 ECTS because empirical evidences suggest that 37 is the best cut-point predictor of university success

4 Results

In Table 1 we assess the balance among the students covariates before and after the matching procedure.

	Unmatched				Matched			
	Treated	Control	SMD	P-value	Treated	Control	SMD	P-value
Male %	71,0%	50,0%	0,44	0,00	66,9%	65,0%	0,04	0,17
Student's ESCS	0,39	0,52	-0,14	0,00	0,48	0,52	-0,05	0,08
HS final mark	78,87	81,56	-0,22	0,00	79,41	79,88	-0,04	0,16
INVALSI math result	245,61	245,20	0,01	0,34	248,83	248,42	0,01	0,66
INVALSI italian result	226,04	233,24	-0,22	0,00	231,05	232,26	-0,04	0,18
INVALSI english result	451,84	459,11	-0,13	0,00	460,38	461,73	-0,02	0,40

Table 1 Student covariates comparison in the treatment and control groups before and after the matching.

According to the standardized mean difference (SMD) [5] before the matching procedure, the major imbalance is in terms of gender, while no imbalance is recorded for the INVALSI math results. After the matching procedure all the student covariates are balanced, in fact no SMD is statistically significant.

In Table 2, the results of two logistic models with random effects on the university field of study fitted on unmatched and matched data are reported. The main differences are: before matching, the treatment effect was negative and significant, suggesting that the applied sciences students have worse performance with respect to their peers of the traditional scientific; after matching, the treatment is not anymore effective. Moreover, in Figure 1, it is possible to see how the treatment varies after the matching procedure with respect to the field of study: if on unmatched data the treatment had a positive effect for the STEM courses and negative on no STEM courses, after the matching the effect is reversed.

5 Conclusions

In this paper we compare the university performance of students that attended two different school tracks, to assess the effect of different skills. The main results show that the two groups significantly differ by gender, socio-economic level, final high school grade and scores in English and Italian INVALSI tests, while no difference are observed in maths test scores. After having balanced both school and students

Variable	Unmatched		Matched	
	Estimate	Pvalue	Estimate	Pvalue
Intercept	-5,48	0,00	-5,98	0,00
HS diploma (ref="Scientific")	-0,11	0,05	-0,04	0,56
HS final mark	0,07	0,00	0,07	0,00
Genere (ref="Female")	-0,14	0,00	-0,22	0,01
HS macro-region (ref="North") - Center	-0,61	0,00	-0,59	0,00
HS macro-region (ref="North") - South and Islands	-0,98	0,00	-0,92	0,00
Student's ESCS	0,06	0,00	0,04	0,26
INVALSI math result	0,01	0,00	0,01	0,00
INVALSI italian result	0,00	0,95	0,00	0,16
INVALSI english result	0,00	0,00	0,00	0,54

Table 2 Parameters of logit regression models with random effects before and after matching.

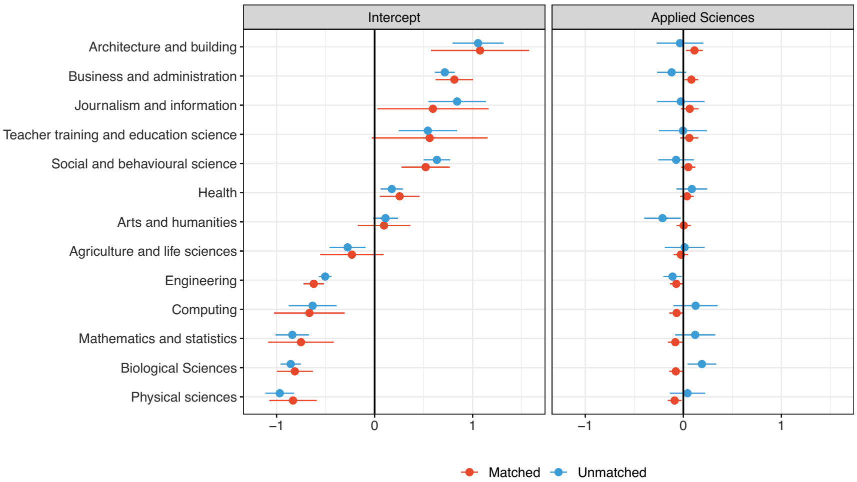

Fig. 1 Comparison of the intercept and the treatment (Applied Sciences) parameters for each university field of study estimated using both the unmatched and the matched data.

covariates, the results of a logit regression model with random effects with respect to the university field of study suggest that there is no difference between the two tracks. However, some considerations are needed. We evaluate the difference among the two different school tracks just with respect to their academic performance. In reality, it already exists a difference in the enrollment rates of the students belonging to the two groups, strongly in favor of the traditional scientific school track. This tunes the matching approach towards a specific direction: it selects the best students of the applied sciences school track. The Italian high school system is still hierarchical and the traditional scientific track is then considered as characterized by a better "social context" compared to the applied science track. Our procedure can not balance for this social context, which is an unobservable variable.

Acknowledgements This paper has been supported from Italian Ministerial grant PRIN 2017 “From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North-South divide.”, n. 2017HBTK5P.

References

1. Aina, Carmen, Massimiliano Bratti, and Enrico Lippo. ”Ranking high schools using university student performance in Italy.” Economia Politica 38.1 (2021): 293-321.
2. Azzolini, Davide, and Loris Vergolini. ”Tracking, inequality and education policy. Looking for a recipe for the Italian case.” Scuola democratica 2 (2014): 0-0.
3. Cascella, Clelia, Chiara Giberti, and Giorgio Bolondi. ”An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks.” Studies in Educational Evaluation 64 (2020): 100819.
4. Checchi, Daniele, and Luca Flabbi. ”Intergenerational mobility and schooling decisions in Germany and Italy: The impact of secondary school tracks.” (2007).
5. Cohen, Jacob. ”Statistical power analysis for the social sciences.” Hillsdale, NJ: Lawrence Erlbaum (1988).
6. Contini, Dalit, Di Tommaso, Maria L., and Mendolia, Silvia. ”The gender gap in mathematics achievement: Evidence from Italian data.” Economics of Education Review 58 (2017): 32-42.
7. D’Amore, Bruno, et al. ”La didattica della matematica: strumenti per capire e per intervenire.” Atti del Convegno Nazionale omonimo (2014): 3-4.
8. Hagan, John Ekow, et al. ”Students’ perception towards mathematics and its effects on academic performance.” Asian Journal of Education and Social Studies 8.1 (2020): 8-14.
9. Keele, Luke, Matthew Lenard, and Lindsay Page. ”Matching Methods for Clustered Observational Studies in Education.” Journal of Research on Educational Effectiveness 14.3 (2021): 696-725.
10. Nicholas, Jackie, et al. ”Mathematics preparation for university: entry, pathways and impact on performance in first year science and mathematics subjects.” International Journal of Innovation in Science and Mathematics Education 23.1 (2015).
11. Panichella, Nazareno, and Moris Triventi. ”Social inequalities in the choice of secondary school: Long-term trends during educational expansion and reforms in Italy.” European Societies 16.5 (2014): 666-693.
12. Petolicchio, Annamaria. ”La Riforma del Liceo Scientifico: il curricolo Scienze Applicate.” FORMAZIONE & INSEGNAMENTO. Rivista internazionale di Scienze dell’educazione e della formazione 14.2 (2016): 187-194.
13. Pimentel, Samuel D., et al. ”Optimal multilevel matching using network flows: An application to a summer reading intervention.” The Annals of Applied Statistics 12.3 (2018): 1479-1505.
14. Ricci, Roberto. ”The Economic, Social, and Cultural Background: a continuous index for the Italian Students of the fifth grade.” Mimeo (2010).
15. Rosenbaum, Paul R. ”Modern algorithms for matching in observational studies.” Annual Review of Statistics and Its Application 7 (2020): 143-176.
16. Xie, Yu, Michael Fang, and Kimberlee Shauman. ”STEM education.” Annual review of sociology 41 (2015): 331-357.