

The consequences of promoting data literacy among non-Stem graduates

Susanna Tinti (Università di Bologna); Margherita Fort (Università di Bologna); Annalisa Loviglio (Università di Bologna)

Research question: The paper looks at the effects of a training program targeting transferable skills - specifically data literacy- at graduate level. Primary outcomes are data literacy competencies but also the effects on students' university careers.

Institutional setting: The paper considers a case study in which a training is offered to students enrolled in master classes in a large Italian University. Participation is voluntary and allows students to get formal qualifications complementary to their master's degree, by attending 4 additional classes. The training unfolds over about 6 months for a total of about 140 hours of classes. The set of 4 courses aims at enhancing analytical skills, and targets students whose academic backgrounds have minimal exposure to quantitative subjects.

The program was designed by scholars in the fields of Statistical Science, Computer Science and Engineering, Economics and Management, with the goal of offering a multidisciplinary approach to data analysis and interpretation.

Methods: As the number of applications exceeded the number of available slots for the program, rationing was resolved through random assignment. Applicants were randomly assigned to either a group of students offered the possibility to enroll in the program, or a group not- offered the possibility to enroll. Waiting lists were also created based on the results of the random assignment. We observe two-sided non-compliance and we can thus assess the reduced form causal effect of being offered a slot in the program on participation and outcomes, as well as the causal effect of participating in the program on outcomes, relying on an instrumental variable strategy, that exploits random assignment for identification.

Data: We use data from the academic year 2021-2022, when the program was introduced for the first time. Our empirical analysis is based on both administrative and survey data. We use administrative data to collect pre-treatment information on students, attendance, and performance of the four classes offered within the program, plus post treatment information on students' main university career.

The background data information cover: i) demographics such as gender and nationality; ii) university

career, such as course they are enrolled in, home campus, exams passed and grades. In addition, we also collect pre-treatment information on the levels of numeracy and literacy based on validated standardized tests (Programme for the International Assessment of Adult Competencies – PIAAC). We designed and administered online an ad hoc survey, in the same spirit of GMAT and GRE validated test, aimed at measuring data literacy. The objective of the questionnaire was evaluating students' proficiency in applying logical-mathematical reasoning to address and solving real-world problems and their capacity of understanding and interpreting tables and graphs.

Results: Endline response rates were around 80% and well-balanced across individuals offered a slot in the program and individuals not-offered the position. Both groups are balanced also in terms of observed pre-treatment characteristics, including numeracy and literacy as measured by PIAAC test.

Preliminary results suggests that individuals participating in the program and starting from low-numeracy level benefit from the program, while at the same time no negative effect is observed on their career path at university (namely number of exams passed).

Limits: We acknowledge that our study is based on limited sample size and results should be interpreted with some caution.

Contribution and next steps: Results of this paper suggest that interventions targeting individuals with expected low levels of data literacy may be successful in improving the targeted transferable skill. Further research is needed to explore the effect on their employability and potential positive externalities for the society as a whole, as citizens with higher data literacy skills might be able to make better informed decisions.